POLITECNICO DI TORINO
Repository ISTITUZIONALE

Seamless Positioning and Navigation in Urban Environment

Original
Seamless Positioning and Navigation in Urban Environment / DI PIETRA, Vincenzo. - (2019 Apr 18), pp. 1-247.

Availability:
This version is available at: 11583/2732878 since: 2019-05-10T09:55:50Z

Publisher:
Politecnico di Torino

Published
DOI:

Terms of use:
Altro tipo di accesso

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

04 May 2023



Doctoral Dissertation
Doctoral Program in Urban and Regional Development (31.st cycle)

Seamless Positioning and Navigation in Urban Environment

Vincenzo Di Pietra
* * * * * *

Supervisors
Prof. Andera M. Lingua, Supervisor
Prof. Marco Piras, Co-supervisor

Doctoral Examination Committee:
Prof. Sanna Kaasalainen, Referee, Finnish Geospatial Research Institute
Prof. Antonio Vettore, Referee, Università degli Studi di Padova
Prof. Laura Ruotsalainen, University of Helsinki
Prof. Domenico Visintini, Università degli Studi di Udine
Prof. Stefano Gandolfi, Università di Bologna

Politecnico di Torino
2019

This thesis is licensed under a Creative Commons License, Attribution - Non commercial- NoDerivative Works 4.0 International: see www.creativecommons.org. The text may be reproduced for non-commercial purposes, provided that credit is given to the original author.

I hereby declare that, the contents and organisation of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

. . . . . .

Vincenzo Di Pietra

Turin, 2019

Summary

In the era of Information Society, where people, objects, spaces and vehicles are rich of sensors able to acquire bites of data and to share them through an interconnected communication network to provide numerous services, it is fundamental to link these information and these sensors with their global position in the world. Location Based Services (LBSs) consist in a widespread panorama of applications, tasks, and services in which the position is the fundamental information to provide. Some examples are search and rescue operations, autonomous navigation, impaired assistance, geospatial analysis, Intelligent Transport System (ITS) and behaviour characterization. This wide panorama of applications requires robust and reliable estimation of the position, both indoor and outdoor, and in the transition between them. Since sensors performances are strictly bound to the type of environment, the research effort must be driven to the ubiquity of the solution which means integrating different sensors to reach the seamless positioning and navigation.

Nowadays, the location information is usually provided by a constellation of satellites that communicate with a receiver on the earth i.e. the Global Navigation Satellite System (GNSS). Thanks to this technology it is possible to estimate the global position, velocity and time of the receiver with high accuracy. Unfortunately, as people spend most of their time indoor and in urban areas, their positioning devices experience huge limitation. Indoor, the GNSS signals are attenuated or blocked by objects in line of sight and the positioning is highly degraded or unavailable. Moreover, urban environment presents some challenges that make GNSS vulnerable to a range of threats: multipath caused by the presence of buildings and trees, decreased satellite visibility and received signal strength, due to no open-sky condition, interferences from anthropogenic radiofrequency emissions over the GNSS bandwidths and more. Overcoming all these issues requires methods, technologies and sensors able to enhance the system performances in term of accuracy, reliability and continuity, both indoor and outdoor. For this purpose, multi-sensor integration and data fusion must be adopted to exploit the benefits of several complementary technologies and to limit their weaknesses. Moreover, the market demands for ubiquitous solutions embedded in low-cost devices. This requirement introduces further challenges related to the noise affecting low-cost sensors and to the seamlessness of the solution, i.e. the continuity of the navigation

estimation in the passage from outdoor to indoor spaces and vice versa. In this panorama, the present work aims to investigate on low-cost sensors integration for positioning and navigation in challenging environments. Based on the literature review and focusing the interest on the urban environment, the most suitable sensors have been selected, tested and validated in different scenarios. The characterization of these systems (i.e. technology, methodology, algorithms) has been used to implement an hybridized navigation solution which integrates Ultra-wideband (UWB), visual sensors, and Inertial Navigation System (INS). The results of this integration have been applied in one of the most challenging environment: an underground mine.

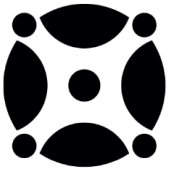
Acknowledgements

The research presented in this thesis has been carried out during last three years in the Department of Environment, Land and Infrastructure Engineering (DIATI) of Politecnico di Torino. During this period I had the privilege to work with numerous people, colleagues of the Goematic Research Group, who have contributed to my work and have made my everyday work pleasant. I would like to express them a special thanks for all the fun we have had during these years.

I would like to thank my supervisor Prof. Andrea Maria Lingua and my co-supervisor Prof. Marco Piras for providing me the opportunity to develop this research and to expand my scientific knowledge.

Also, I have been privileged to visit the Department of Navigation and Positioning at the Finnish Geospatial Research Institute (FGI) in 2018. I would like to express my appreciation to Prof. Laura Ruotsalainen for the opportunity to work in Sensor and Indoor Navigation (SINA) group and to Prof. Sanna Kaasalainen for involving me in Kaivos Project. A special thanks goes to my friends in FGI.

My research have been supported financially by Compagnia di San Paolo during the Urban and Regional Development Phd program activated by Politecnico di Torino and Università degli Studi di Torino. In addition, I would like to acknowledge the PoliTO Interdepartmental Centre for Service Robotics (PIC4SeR).



PIC4SeR
PoliTO Interdipartimental Centre for Service Robotics

Torino, March 2018
Vincenzo Di Pietra

I would like to dedicate this thesis to my loving family, Vanessa, Samuele, Giuseppe and Angela; and to my two beautiful nephews, Edorardo and Alessandro, who have made this journey lighter and joyful.
My greatest thanks go to Corinna, who has supported me during these years with her smile.

- “Vincent! How are you doing this Vincent? How have you done any of this? We have to go back!”

- “It’s too late for that...We’re closer to the other side.”

- “What other side? You wanna drown us both”

- “You wanna know how I did it? This is how I did it Anton. I never saved anything for the swim back..”

Gattaca, 1997

Contents

List of Tables

List of Figures

List of Symbols

Roman Symbols

Δf	Frequency bandwidth
η	Image coordinate y-direction
κ	Heading or distortion parameters
λ	Carrier wavelength or longitudeor lapse rate
ω	Earth turn rate
ω	Roll
ω_{ib}^{b}	Body turn rate with respect to the inertial frame angular velocity measurement
ϕ	Carrier-phase or Latitude
ρ	Range measurement
φ	Pitch
$\widehat{\omega}$	Gyroscope angular velocity measurement
$\widehat{f}$	Specific force
ξ	Image coordinate x-direction
c	Speed of light
C_{b}^{n}	Rotation of the body frame to the navigation frame
d	Object distance from camera center
E	Essential matrix or ephemeris error
F	Fundamental matrix

f	focal length
g	Gravity vector
H	Design matrix or image homography
I_k	Image matrix at k time
M	Multipath error
n	Normal to plane
P	State error covariance or camera matrix or Pseudorange
P_0	Pressure at sea level
P_m	Measured pressure
R	Measurement noise covariance or camera rotation matrix or universal gas constant
rr	Range resolution
T	Tropospheric propagation
t	time
u	Principal point's x-coordinate
v	Principal point's y-coordinate
v_k	State noise vector
w_k	Measurement noise vector
x	State vector
y	Measurement vector
$\hat{h}_B$	Relative altitude

List of Abbreviations

AoA	Angle of Arrival
A-GPS	Assited GPS
A-GNSS	Assisted GNSS
CORS	Continuously Operating Reference Station
CDF	Cumulative Density Function
EKF	Extended Kalman Filter
FCC	Federal Communications Commission
FGI	Finnish Geospatial Research Institute
GNSS	Global Navigation Satellite System
GDOP	Geometric Diluition Of Precision
HSL	Hyperspectral LiDAR
INS	Inertial Navigation System
IMU	Inertial Measurement Unit
IPS	Indoor Positioning System
IRBL	Image Recognition Based Location
ITS	Intelligent Transport System
KF	Kalman Filter
LSE	Least Square Estimation
LSR	Least Square Residual
LBS	Location Based Service

LOS	Line of Sight
LiDAR	Light Detection and Ranging
MEMS	Micro Electro-Mechanical System
NLLS	Non Linear Least Square
NLLSE	Non Linear Least Square Error
NLOS	Non Line of Sight
OS	Operating System
PDF	Probability Density Function
RFID	Radio-Frequency IDentification
RANSAC	RANdom SAmple Consensus
RTK	Real Time Kinematic
RSSI	Received Signal Strength Indicator
SINA	Sensor and Indoor Navigation
SNR	Signal to Noise Ratio
SOC	System of Chipset
SURF	Speeded up robust features
ToA	Time of Arrival
TDoA	Time Difference of Arrival
ToF	Time of Flight
TWR	Two Way Ranging
UWB	Ultra-wideband
UWB-IR	impulse radio Ultra-wideband
WLSE	Weighted Least Square Estimation

Chapter 1

INTRODUCTION

Locating our position and moving in an unknown neighbourhood are relatively simple activities we execute making use of our sight and memory. This is true when we perform simple tasks in familiar environment such as finding an exit in a public infrastructure or locating the stairs to go up one floor. When the task is more challenging, like finding the route to a certain place or piloting an intercontinental flight, our memory is no longer useful and we need to estimate our position with external data. In order to do this, several types of data are required, such as geometrical, semantic and topological information of the surrounding space. We need sensors able to capture data from the external environment allowing us to respond properly to inputs, in view of different purposes. Moreover, when we want to know the location of objects or vehicles we acquire data transmitted from these objects and we estimate their position in relation with our own. Is interesting to observe how, for the world population, knowing always their own position is an obvious task, although the estimation procedures are complex.

With the appearance of Global satellite Positioning Systems (GPS), the positioning and navigation operations have become common features, necessary for the conduction of everyone's daily activities. The development of GPS technology enabled to locate objects, means of transport, people and to track and direct their movements, assuring accurate data exchange, excellent performances and proper responses in term of safety. Still, if satellites-based positioning and navigation technologies has reached a sufficient degree of development and widespread availability outdoor, equally efficient solutions are needed when positioning data fails caused to low satellite view. Furthermore, assessing the GNSS signal integrity in complex urban environment is not easy, not only for the visibility of the constellation but also for other challenges like electromagnetic interferences, atmospherics factors, reflective surface and more. Moreover, as people live in a continuous transition from indoor spaces to open areas, these solutions need to be seamless and able to provide location information both indoor and outdoor. In this context raised the interest of the academic community to investigate and develop methods and

techniques to obtain positioning, navigation and tracking estimation with higher level of accuracy from different data and sensors integrated in a unique solution. A multi-sensor positioning system that allows the data continuity and integrity in the passage from external environments to indoor environments and vice versa. The urban areas represent a meaningful environment in which test the proposed solutions and develop innovative applications.

The plethora of sensors integrated in seamless positioning systems are based on electromagnetic and mechanic waves measurements or inertial navigation. There are imaging sensor acquiring visible and infrared light. Inertial navigation systems acquiring accelerations and angular rotation of a body. Ultrasound-based sensors and radiofrequency systems including radars, cellular networks, wireless sensor networks, Radio-Frequency IDentification (RFID), Bluetooth and Ultra-wide band. When a single or an integrated technology is implemented for a particular application meeting users requirements, arises LBS which has become increasingly important in a large number of applications in recent years.

1.1 Research objectives

In this thesis the focus has been pointed on the environment, typical scenarios like indoor offices and underground infrastructures where the positioning is particularly challenging. Although numerous researchers have investigated this topic, solutions able to face the problem of ubiquitous positioning no exist at the current state of the art. Moreover, low-cost sensors causes further challenges in solve estimation procedures and requires additional effort. The proposed solutions, based on multi-sensor integration algorithms, are able to perform robustly and accurately in particular environment with a controlled context. Interferences, kinematic condition, scalability, working rate becomes site-dependent variables that prevent to reach the objectives of seamless navigation.

In order to identify the most suitable technologies and methods to perform multi-sensor positioning and navigation in urban environment, the most common self-contained sensors used in pedestrian navigation has been tested in real case studies. The acquired data have been analysed in term of performances and behaviour characterization in relation with the surrounding. In particular, after a deep analysis on the state of the art, the sensors has been compared and evaluated in function of different parameters like accuracy, integrity, cost, energy consumption and more. The panorama is huge and the main technologies has been elected according to the following requirements. Any urban city can be divided in function of the characteristic of the surrounding environment. One characteristic is the size in which users or vehicles should operate and navigate. A room or a building floor represent different environmental characteristics with respect to an infrastructure or an open city park. Moreover, the operational scale is also one of the parameters

of any positioning system and define its capability to provide position and navigation information within a certain area. Trying to give a structure to this concept, the urban environment has been classified in Floor scale spaces, Infrastructure scale spaces and District scale spaces.

Therefore, the UWB technology has been chosen to perform positioning in floor scale environment, the vision-based positioning and navigation has been elected as best technology for an infrastructure-like building and the new-generation GNSS smartphone chipset has been used in the district scale spaces.

These three technologies has been used, analysed and validated in real urban areas corresponding to the characteristics previously described, and algorithms and methods have been developed to estimate the user location. These tests allowed to define limits and advantages of these sensors, to characterize their behaviour in terms of accuracy, integrity, noise, and finally to propose a navigation system. The proposed solution, composed by an integration of low-cost sensors and a developed fusion algorithm, has been tested in one of the most challenging environment: an underground mine.

1.2 Motivation

According to the World Urbanization Prospects 2018 provided by the Department of Economic and Social Affairs of the United Nations by 2050, the world population is expected to reach 9.6 billion, increasing by 68% the inhabitants of the urban areas. In Europe, the 80% of the total population is expected to live in cities and this will generate profound changes in the urbanization process from the social-economic, climatic and environmental point of view. The disaster risk, the security, the safeguard of the cultural heritage, the vulnerability of the historical centers, the social innovation and the needs of the citizen to taking part to the decision-policies, are fundamental topics that need to be investigated.

With sensors becoming part of our buildings, transportation systems and personal communication devices, urban areas are increasingly producing digital data of the surrounding environment, flooding the urban digital space of geo-referenced information. In this scenario, providing innovative methods and approaches trough technological tools, which use all these data and allow to better understand the world, becomes essential. The location-aware computing, i.e. the ability of a technological system to define the position of objects and persons, offers innovative tools in supporting policy-makers with their emergency tasks, as well as first responders in the post-disaster scenarios, and citizen with the learning of the urban spaces. Not only, these systems could also be useful as navigation tools or awareness devices. In this context, positioning technologies can make the difference in managing emergency situations in large buildings and underground structure where the possibility to localize and track objects and people is compromised. For example, being aware

of the team location can increase safety and decrease mission time in a first response scenario; moreover, a navigation system reduce the possibility of disorientation and failure to locate victims. Also the same victim could collect information and take action under the possibility to orient himself. As the world continues to urbanize, sustainable development depends increasingly on the successful management of urban activities and therefore on development of smart tools able to rely on accurate and robust location. In last years, smart mobility applications and ITS are becoming also relevant especially for those urban agglomeration with million inhabitants. Safety, security, advertising, health, monitoring and management are few examples of applications have been object of study in the last years [1]. For instance, tracking position of medicines and workers into an hospital can be significantly relevant in order to act in emergencies [2]. In case of fire, locating victims and firemen can improve rescue operations [3]. People suffering from visual, auditory and cerebral deficits could move more autonomously in closed and crowded environments [4], etc. Nowadays, problems related to indoor positioning or indoor navigation have found solutions with the development of several applications conceived for specific buildings as airports or commercial centres. These pioneering experiments, combined with the increasing interest by the industries, are opening to new scenarios and opportunities for applications assisting users in their activities.

Beyond the large number of applications for private and public interest, recent widespread of low-cost mobile devices and ceaseless development of hardware technologies and computational algorithms raised interest from scientific community, which faced the indoor positioning problems using different approaches and technologies. In the last few years common technologies includes sensors related to images, infrared rays, radiofrequencies, ultrasounds, inertial measurements, magnetic fields and atmospheric pressure.

Ten-year research efforts in developing indoor positioning products and services made clear that different solutions are available and they can be based on a combination of technologies. Each system reveals advantages and limits which make it more or less convenient according to the final application. Consequently, the additional effort requested from the scientific community is experimenting integrated systems which compensate limits typical of such technology with advantages deriving from another [5].

1.3 Applications of LBS

The importance of positioning capabilities in our modern way of life is supported by the numerous applications of services and researches based on the location.

Looking at the urban environment, the first relevant applications have been developed for commercial purposes. Mass market applications, which make use of the geographical position, deliver context-dependent information accessible with

mobile device. For example, providing safety information or topical information on cinemas, concerts or events in the vicinity. LBS applications include navigation to the right store in a mall. Products detection, location-based advertisements, location-based billing and local search services have a high commercial value. Applications in train or bus stations include the navigation to the right platform or bus stop. Further examples of LBS are proximity-based notification, profile matching and the implementation of automated logon/logoff procedures in companies. There is also added value for the positioning provider, e.g. by resource tracking, fleet management and user statistics.

Context detection and situational awareness are also fundamental services for the citizens of a large city. Personal mobile devices can provides large variety of useful functions when it is desirable to have an automated adaptation of the mobile device depending on a change of the users context. Such functionalities spares the user additional effort by providing assistance in individual situations. To enable such an automatic adaptation the users context needs to be determined by the mobile device itself. The most significant criteria to determine the users context is the current geographical location. For example a smart conference guide can provide information about the topic discussed in nearby auditoriums.

In hospitals the location tracking of medical personnel in emergency situations has become increasingly important. Medical applications in hospital also include patient and equipment tracking, e.g. fall detection of patients. Precise positioning is required for robotic assistance during surgeries [6].

Observing some phenomenon such as heat, pressure, humidity, air pollution and deformation of objects and structures is fundamental for environmental monitoring. To this end multiple sensor nodes can be located in the environment and connected to a communication network. The data, acquired in real time and correctly georeferenced can be used by policy makers and researchers.

Positioning capabilities, both outdoor and indoor, provide important benefits in law enforcement, rescue services, and fire services, i.e. location detection of firemen in a building on fire. The police benefits from several relevant applications, such as instantaneous detection of thieves or burglaries, detection of the location of police dogs trained to find explosives in a building, detection and recovery of stolen products for post-incident investigations, crime scene recovery, statistics and training but also crime prevention, e.g. with tagged devices for establishing so-called geofencing i.e. alarm systems which can detect whether a person or an asset has left a certain unauthorized area.

Intelligent transportation is another emerging field thanks to the users expectations for greener, safer and leaner mobility. Autonomous driving is the leading aims but mass user application for indoor navigation in parking are already experimented.

Several applications have been developed in museums, such as visitor behaviour

characterization, visitor tracking for surveillance, location based guiding and triggered based on context aware information.

Structural health monitoring consist in sensors incorporated into steel reinforcements within concrete to perform strain measurements with high resolution. Strain sensing systems based on passive sensor and integrated RFIDs can measure strain changes and deformation caused by loading and deterioration.

Strict positioning requirements are needed in tunneling [7] and longwall [8] where dusty, darkness, humid and space limited environments generates more challenges.

Scene modelling and mapping, i.e. the task of building digital 3D models of natural scenes,requires the precise orientation of the optical sensor. Indoor mapping systems need to know the camera position in order to merge multiple views and generate 3D point clouds. Scene modelling is beneficial for several applications such as computer animation, notably virtual training, geometric modelling for physical simulation, mapping of hazardous sites and cultural heritage preservation.

1.4 Author's contribution

The main contributions of the thesis are:

- Provide a comprehensive overview of the GNSS positioning in outdoor environment, along with the wide panorama of Indoor Positioning System (IPS) and their hybridization.
- Define systems requirements based on different applications and meaningful scenarios.
- Select the core technologies in function of three operational scale (floor scale, infrastructure scale and district scale) and assess their performances in real environments.
- Contribute in research and develop of an indoor vision-based positioning system implemented in several Korean sub-way stations.
- Develop of one of the first RINEX parser for smartphone GNSS raw measurements.
- Study of the feasibility and applicability of Loosely Coupling integration algorithm between GNSS, UWB, INS and visual sensor and propose of a framework for low-cost seamless pedestrian navigation.
- Test and analysis of the proposed solution in a challenging environment: an underground tunnel in Otaniemi - Finland.

The research made during the Ph.D. and the work presented in this thesis have led to some peer-reviewed journal papers [9], [10] and to some proceedings in different national [11] and international conferences [12] [13]. Furthermore a book chapter have been published in [14].

1.5 Structure of the thesis

In Chapter 2, "Background on seamless navigation", the definition of the positioning and navigation both in indoor spaces and in seamless condition is given. Then the estimation problem is presented and the principles and algorithms for pose estimation, outlier rejection and quality control are described. The prevalent methods and technologies for navigation are presented in Chapter 3 from the theoretical pointy of view. In particular, the GNSS positioning with smartphones, the image based positioning, the ultra-wide band networks and the integration of GNSS and INS in a loosely coupling approach. Chapter 4 presents methodology, test performed and results of the previous technology in three different real case study. A floor scale test for the UWB positioning, an infrastructure scale test for the image-based positioning and a district scale test for the GNSS positioning with smartphones. In Chapter 5 a proposed navigation solution for underground mines is presented. The data acquisition, the sensors characteristics, the measurements integration and the fusion algorithm are deeply described together with the results of the experiments. The feasibility is demonstrated. Chapter 6 provides conclusions and recommendations for future research.

Chapter 2

BACKGROUND ON SEAMLESS NAVIGATION

In the literature, the process of determining the position and the movement of a body over the time have no universally agreed definition. Numerous technical therms are used to express the same concept or concepts with slightly different meanings. For example, "positioning", "localization", "geolocation", "location sensing" and more, are all terms used in different research fields to express the same estimation problem.

In general, **positioning** is the capability to define the location of a person, an object, or a mean of transportation in a given time with respect to a reference grid or system.

In [15] **navigation** is defined as "...the determination of the position and velocity of a moving body with respect to a known reference". In general, is possible to define navigation as the set of methods and techniques that allow the determination of one or more states (position, speed and attitude) of an object in motion in a predetermined reference system. These states allow the determination of the trajectory made by a body from one point to another of the space. Coordinates, angles, translation vectors, directions, etc. are information which express the states of the body in relative or absolute systems. This state is estimated from a set of measurements acquired by different sensors and related by linear or non linear dependences. Absolute positioning refers to a global reference system realized with a constellation of satellites, allowing to define position in geographic or cartographic coordinates. Relative positioning systems are based on the computation of the position or direction starting from a point with known coordinates. The term "localization sensing" is used in computer science to express information about the location of devices by employing the internet.

Another term used to describe the process of repeated positioning of a moving object over the time is "tracking". In contrast with navigation, tracking is used when the infrastructure is determining the location of a passive mobile device,

where the current position is not necessarily known at the mobile device. In [2], Mautz provides a complete list of prevalent terms for positioning and navigation.

GNSS is the main technologies for positioning and navigation. It is based on the computation of the distance between a GNSS receiver on the earth and various constellation of artificial satellites in known orbit. The distance computation is based on radio frequency transmission; if an obstacle shields the signal, the measurement is significantly degraded and the position is no more available. This is the typical situation of the indoor spaces and urban environment, where the presence of obstacles like buildings, trees and walls generates heavy multipath and attenuation of the GNSS signal. Research on alternative technologies and methods may produce more efficient algorithms, sensor systems and applications in this challenging environment.

Indoor positioning defines the process that attempt to provide an accurate position in a covered structure where the GNSS technology is ineffective. The presence of an infrastructure allows the use of pre-installed sensor networks in the environment providing the position of a target sensor. Moreover, the possibility to inject map information, like floor plan from architectural CAD drawings, enhance the possible positioning solutions. Another scenario is the urban one (or the so-called urban canyons), where the presence of high surfaces such as tall buildings and trees generates multipath effects in the transmission of satellite signals and the GNSS positioning is lost. In this case, other types of active and passive sensors may be integrated with the measurements obtained from the GNSS receiver. In this scenario, the present work aims to investigate on methodologies, technologies and integrated systems which guarantee the positioning and navigation solutions in both outdoor and indoor environments; a solution robust to the vulnerabilities of the GNSS, which works regardless to the environment (i.e. anytime/anywhere). This solution is known under the definition of "seamless positioning and navigation".

2.1 Classification of Indoor Positioning Systems

In previous paragraph, the indoor positioning definition has been given in contrast with the outdoor GNSS positioning. This duality does not express the wide and heterogeneous panorama of the present research field. close spaces and outdoor environments are not separate boxes but a continuum of the urban spaces with interfaces areas where people move and live. Unfortunately, most of the effort in the academic research was given to the investigation of indoor positioning and navigation. It is clear that the research on seamless navigation need to start from the knowledge of the indoor positioning systems.

There are a certain numbers of prevalent technical terms for positioning in literature and all of them define a process of localization that could greatly change according to the type of technology used, to the ultimate aim of the location and

to the type of measurements performed. An initial survey on the literature concerning IPSs has revealed an heterogeneous systems panorama, in which the type of sensors, the location technology, the physical quantity measured, etc. generate different classification methods and different hierarchical levels between systems. Other factors, like the ranging methods and the positioning-navigation technics, characterize the indoor positioning field [16].

Daek et al. in [17] classifies the localization processes in active and passive systems where "active" means that the tracked persons or objects participate actively through an electronic device which sends information to a positioning system helping it to infer that people position. In the passive localization case, the position of the person is estimated on the variation of a measured signal or video process. Thus, the tracked person is not carrying any electronic devices to infer the user's position (Figure 2.1).

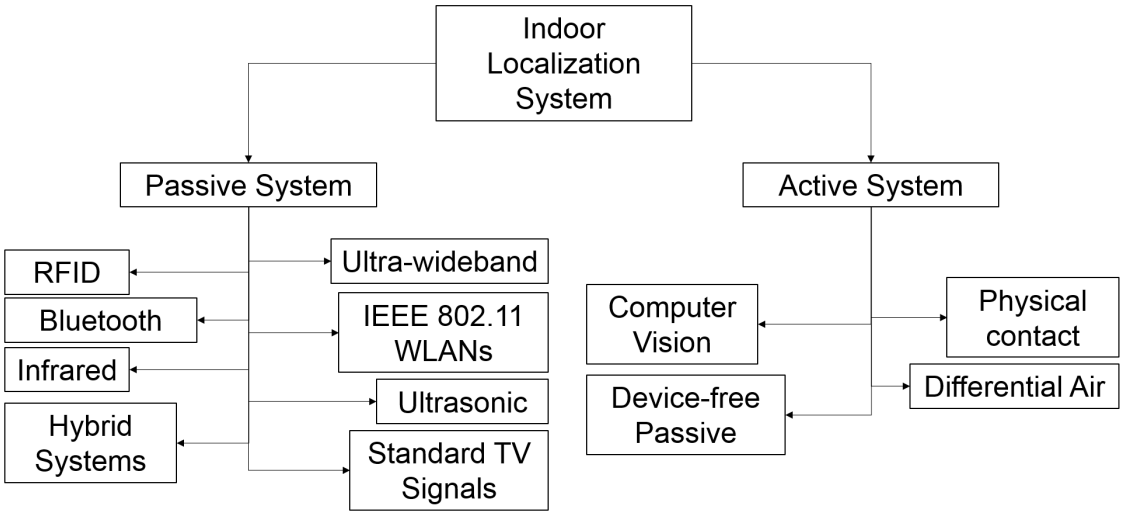


Figure 2.1: Classification of active and passive indoor localization systems.

Jorge Torres-Solis et al., in their *Review of indoor localization technologies* [4] propose a classification of these technologies according to physical quantity measured in the process of localization. In particular, radio frequency waves, photonic energy, sonic waves, mechanical energy, magnetic fields and atmospheric pressure are the quantity grouping the technologies (Figure 2.2).

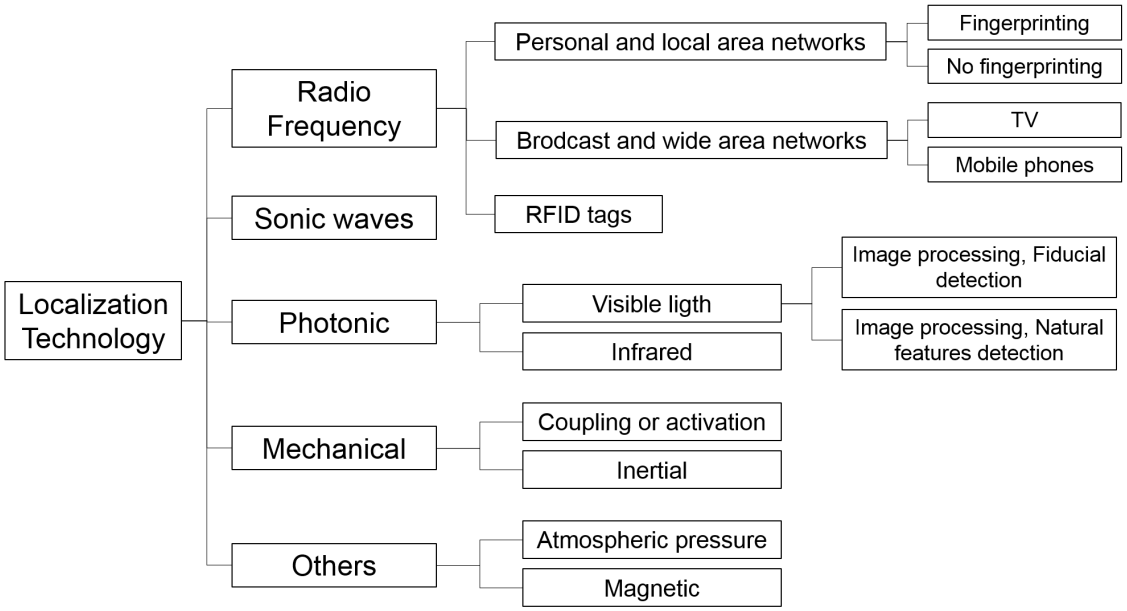


Figure 2.2: Classification of indoor positioning according to the physic quantity measured.

The physical quantity that allows the location must be measured by a sensor, a measuring instrument, or a technology and this enables a further classification of indoor positioning systems. The different technologies are:

- Cameras
- Infrared
- Tactile and Polar system
- sound
- Wlan/Wi-Fi
- RFID
- Ultra-wideband
- High sensitive GNSSs
- Pseudolites
- Other radio-frequencies
- Inertial Navigation

- Magnetic systems
- Infrastructure-based systems

Camera-based localization systems are methods based on the visual information extracted by the camera sensors. These systems can be classified depending on the specific technique used to perform the localization. In fact, they can be based on the comparison with some a-priori information like reference from 3D or 2D building models [18], reference from images [19], reference from deployed coded target [20]. Among vision systems, there are also systems able to tracking objects, humans or animals directly by a fixed camera system without the necessity of external references [21]. Finally, some systems can incrementally compute their position in real time from a known staring point using the visual odometry algorithms.

The positioning systems based on infrared technologies (IR) are among the most common wireless methods used to localize objects or people through infrared emitters and receivers. These systems can use active beacons, natural infrared radiation or artificial light sources. In [22] the authors use fixed active beacons (infrared receivers) placed in known position in the indoor environment to locate mobile beacons.

Similarly, ultrasound localization systems are based on Time of Arrival (ToA) measurements of ultrasound pulses that travel between the emitter sensor and the receiver. In particular, the object to locate has an emitter sensor whose position is estimate by the measured distances from three or more fixed receive (multilateration technique).

Wi-Fi technologies, as the methods analyzed till now, consist of a system that sends data on networks using electromagnetic waves. The main advantage provided by Wi-Fi, apart from the high rates of transmission, is the ability to provide coverage in a wide range of distance (till 100m). Moreover, the main disadvantage of these networks concerns security and privacy. There are some programs that can capture sent packets and calculate the network password, and thus access it.

RFID systems are based on RFID reader, equipped with an emitter/receiver antenna which interrogates nearby active or passive radiofrequency transceivers (i.e. tags). The tags can be active, i.e., powered by battery, or passive, drawing energy from the incoming radio signal. Usually in these systems, the RFID tags send positon information (by an unique ID) to the RFID reader (also known as RFID scanner) that estimates its position by some procedure of proximity (Cell of Origin) or by the measurement of the Received Signal Strength Indicator (RSSI) [23]. The accuracy of an RFID system is directly related to the density of tag deployment and reading range.

Inertial sensors refer to systems that exploit the inertia to measure linear acceleration or angular velocity (i.e. accelerometer and gyroscope). In this regards, an INS is a navigation aid to use motion sensors and rotation sensors to continuously calculate the position, orientation and velocity (i.e., direction and speed

of movement) of a moving object. Since inertial sensors yield relative positioning information only, an absolute reference is required to specify the displacement reported by an inertial measurement in absolute coordinates. Therefore, to provide absolute positioning information it is necessary to combine inertial sensors with different localization technologies.

2.2 Seamless positioning and navigation

As already state, the seamless navigation can't be achieved using a single technology solution but requires a multi-sensor approach, where the integration of the sensors and the design of the fusion algorithm could benefits of the different behaviour of the measurements. The idea is to integrate already existing solution that perform well in different scenarios, mainly in outdoor and indoor environments. Therefore, derive the need to identify the location context in order to adapt the navigation filter and to switch between sensors. In outdoor, GNSS coupled with other sensors should be preferred over others systems, while indoor should be avoided. Moreover, also the behavioural context is an important information. A sensor used to define the number of steps during a run or a walk, can't be used when the user take the bus. The capability of the system to adapt itself to the surrounding environment is called context awareness or adaptive sensor fusion. Seeing the huge panorama of systems, the choice of the best technology is not an easy task and require a deep investigation on the characteristics of the final application, of the context, of the requirements and about the cost. The main aspects to analyse and requirements to guarantee in order to obtain a seamless system are:

- accuracy
- availability
- robustness
- floor identification
- estimation rate
- scalability
- cost
- maintenance

2.3 Estimation problem in positioning and navigation

The positioning and navigation rely on the estimation of the parameters that define the location of an object, person or vehicle with respect to the local environment or to a global reference frame. The first step to approaching the estimation problem is to define the unknown parameters to estimate i.e. the state vector $\mathbf{x}$. When the objective is to find the position, the unknown are usually the three-dimensional coordinates of the object and his attitudes angles. In trajectory estimation, useful for real time navigation, the parameters are usually the position coordinates in the reference frame, the velocity and the attitudes angles. In GNSS navigation, the unknown are the three Cartesian coordinates plus the clock offset between satellites clocks and receivers clocks.

GNSS positioning	Indoor positioning	Trajectory estimation
$\mathbf{x} = (X, Y, Z, dt)^T$	$\mathbf{x} = (X, Y, Z, \omega, \varphi, \kappa)^T$	$\mathbf{x} = (X, Y, Z, v_x, v_y, v_z, \omega, \varphi, \kappa)^T$

After defining the state vector, the estimation problem consist to known which is the relation between the output of the sensors and how these outputs relate to our known through a mathematical model or so called estimator. Considering this, the problem consist on the definition of the more suitable estimator that put in relation the acquired measurements with the searching unknown. It is important also to be aware that all measurements are effected by errors. Being these errors random variables, also our measurements have to be considered like random variables. Consequentially, the unknown are also randomic and therefore statistic and probability must be applied on the estimation procedure to validate the results. Considering:

state vector that has to be estimated	$\mathbf{x}$
measurement vector modelled by	$\mathbf{y} = \mathbf{h}(\mathbf{x})$
noise vector affecting the measurements	$\mathbf{y} = (\mathbf{h}(\mathbf{x}) + \mathbf{n})$
estimator	$\widehat{\mathbf{x}} = \mathbf{g}(\mathbf{y})$

Estimator are obtained by defining a criteria function $C(x)$ and minimizing this criteria. This criteria function is often related to the estimation error so is minimizing the variance of the estimation error. The estimate value $\widehat{\mathbf{x}}$ is also random therefore one must characterize the quality of this estimate using statistical parameters like the accuracy and the variance.

$$\widehat{\mathbf{x}} = \underset{x}{argmin} C(\mathbf{x})$$

There are two class of techniques of estimation:

- Classical estimation, where no assumption on the parameters to estimate is made. An example is the Least Square Estimation (LSE).
- Bayesian estimation, where is assumed that the variable to estimate is random variable. An example is the Kalman Filter (KF).

In Bayesian estimator, assuming that the estimate parameters are random variables, it is possible to attach to this the Probability Density Function (PDF), for example a Gaussian distribution. This assumption provide more information to design the right minimize cost function and to obtain a more accurate estimator. The problem is that not all the measurements reflect information about the unknown and so classical estimators are usually more robust in this sense. As the estimate value $\widehat{x}$ is a random variable therefore one must characterize the quality of this estimate using statistical parameters like the accuracy and the variance. The accuracy is the discrepancy between the estimated value and the reference one, while the variance or precision is the estimation dispersion with respect to the mean value of the set. In positioning and navigation, these concepts are usually translated in level of confidence, availability, continuity, coverage and reliability.

2.3.1 Linear Least Square Estimation

The least square estimation is a classical estimation technique for overdetermined systems that can be used if there is a linear relationship between unknown and measurements, $y = Hx + n$. The measurements must not be affected by biases and must have gaussian distribution $\varepsilon(n) = 0$. In this case, LSE consist to minimize the Least Square Residual (LSR) i.e. the sum of the squared differences between the modelled measurement vectors and the observed measurements (measurement error) $\Sigma(h(x) - y)^2 = \Sigma v^2$ (in matrix form is $y = Hx; y - Hx = v$). The computation of the minimum value of Σv^2 consist in taking its first derivative and equating the resulting function with zero. From this definition is possible to set a system of linear equations that in matrix form is $Ax = b$ or $Nx = T_n$. with $N = H^T H$ pseudoinverse and $T_n = H^T y$. Then the LSR estimator is:

$$\widehat{\mathbf{x}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{y} \tag{2.1}$$

If the covariance matrix of the noise is known $\mathbf{C}_{yy}$ or $\mathbf{R}$, is possible to predict the covariance matrix of the estimation error and so the covariance matrix of the estimator $\widehat{\mathbf{x}}$:

$$\mathbf{C}_{\widetilde{xx}} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\mathbf{C}_{yy}\mathbf{H}(\mathbf{H}^T\mathbf{H})^{-1} \tag{2.2}$$

On the other hand, if the covariance matrix of the measurement error C_{vv} is known is possible to weight the measurements depending on their variance. In this case the estimation technique is called Weighted Least Square Estimation (WLSE):

$$\widehat{\mathbf{x}} = (\mathbf{H}^T \mathbf{C}_{yy}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_{yy}^{-1} \mathbf{y} \tag{2.3}$$

The covariance matrix of the estimation error $\mathbf{x} - \widehat{\mathbf{x}}$ is:

$$\mathbf{C}_{\widetilde{xx}} = (\mathbf{H}^T \mathbf{C}_{yy}^{-1} \mathbf{H})^{-1} \tag{2.4}$$

with:

$$\mathbf{C}_{yy}^{-1} = \begin{bmatrix} \sigma_1^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_n^2 \end{bmatrix} \tag{2.5}$$

2.3.2 Non-Linear Least Square Estimation

In some cases the measurement model may not be linear, as for example in the case of range measurements. The linear LSE cannot be used directly in this situation but the model must be linearised around a supposed state vector $\widehat{x}_0$. If an approximate value of x is known. Is possible to re-write the function as a first order derivative of Taylor series. The Non Linear Least Square Error (NLLSE) model is:

Non linear measurement model	$\mathbf{y} = \mathbf{h}(\mathbf{x}) + \mathbf{n}$
Linearisation of the model in$\hat{x}_0$	$\mathbf{h}(\mathbf{x}) = \mathbf{h}(\hat{\mathbf{x}_0}) + \frac{\delta \mathbf{h}}{\delta \mathbf{x}}(\widehat{\mathbf{x}}_0)(\mathbf{x} - \widehat{\mathbf{x}}_0)$
Definition of the Jacobian matrix	$\mathbf{H} = \frac{\delta \mathbf{h}}{\delta \mathbf{x}}(\widehat{\mathbf{x}}_0)$
Rewriting of the measurement model	$\mathbf{y} = \mathbf{h}(\hat{\mathbf{x}}_0) + \mathbf{H}(\mathbf{x} - \widehat{\mathbf{x}}_0) + \mathbf{n}$
The transformed measurement	$\Delta\mathbf{y} = \mathbf{y} - \mathbf{h}(\hat{\mathbf{x}}_0)$
The transformed state	$\Delta\mathbf{x} = (\mathbf{x} - \widehat{\mathbf{x}}_0)$
The Least Square Estimation	$\Delta\hat{\mathbf{x}}_{LS} = (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T\Delta\mathbf{y}$
The final estimate	$\hat{\mathbf{x}} = \hat{\mathbf{x}}_0 + \Delta\hat{\mathbf{x}}_{LS}$

Usually this method is applied iteratively in order to refine the linearisation point. The Least Squares estimators are quite easy to implement, and provide generally good results. The main advantage is the lack of assumptions that have to be done in the estimator design.

2.3.3 Kalman Filter

Nowadays, when a navigation system is design, the measurements acquired in real-time are combined with dinamic motion models (velocity, acceleration, etc.) to define the predicted position of the body. The most used approach to perform

this integration is the KF. KF is a recursive estimator means to estimate the state of a process, in a way that minimizes the mean of the squared errors [24]. The set of equations used in KF are based on two estimation models, the state transition model and the measurements model. The recursive nature of the filter provide means of incorporating information about the past state and predict the new state adding new measurements. The assumptions of the KF are the gaussian stochastic model and the linear relation between subsequent state equations:

$$\text{state transition model} \quad \mathbf{x}_{k+1} = \mathbf{F}_k \mathbf{x}_k + \mathbf{w}_k \tag{2.6}$$

$$\text{measurement model} \quad \mathbf{y}_k = \mathbf{H}_k \mathbf{x}_k + \mathbf{v}_k \tag{2.7}$$

where:

- $\mathbf{F}_k$ is the transition matrix
- $\mathbf{w}_k$ is the state noise vector ($\mathbf{C}_{vv}$ or $\mathbf{Q}$) follows a known normal distribution $N(0, Q)$
- $\mathbf{H}_k$ is the observation matrix
- $\mathbf{v}_k$ is the measurement noise vector ($\mathbf{C}_{ww}$ or $\mathbf{R}$) follows a known normal distribution $N(0, \mathbf{R})$

The principle of the Kalman Filter for a dynamic system is shown in Figure 2.3. It is based on two main steps, the state prediction and the measurements update, used in a recursive iterative process. During the prediction, a predicted estimate $\hat{\mathbf{x}}_{k|k-1}$ is obtained by applying the state transition matrix. In the update step, this prediction estimate is corrected by the measurement information.

Prediction:

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}(\hat{\mathbf{x}}_{k-1|k-1}) \tag{2.8}$$

Update:

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k(\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}) \tag{2.9}$$

where $\mathbf{K}$ is the Gain matrix or Kalman matrix. The factor $(\mathbf{y}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1})$ inside the brackets is called innovation state or predicted residual and represent how much the new measurements $\mathbf{y}_k$ are "good" for the predicted estimator $\hat{\mathbf{x}}_{k-1}$. The $\mathbf{K}$ matrix is computed as:

$$\mathbf{K}_k = \mathbf{\Sigma}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{\Sigma}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k)^{-1} \tag{2.10}$$

For both step is possible to compute the state covariance prediction 2.11 and the state covariance update 2.12:

$$\mathbf{\Sigma}_{k|k-1} = \mathbf{F}_k \mathbf{\Sigma}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_k \tag{2.11}$$

$$\boldsymbol{\Sigma}_{k|k} = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\boldsymbol{\Sigma}_{k|k-1} \tag{2.12}$$

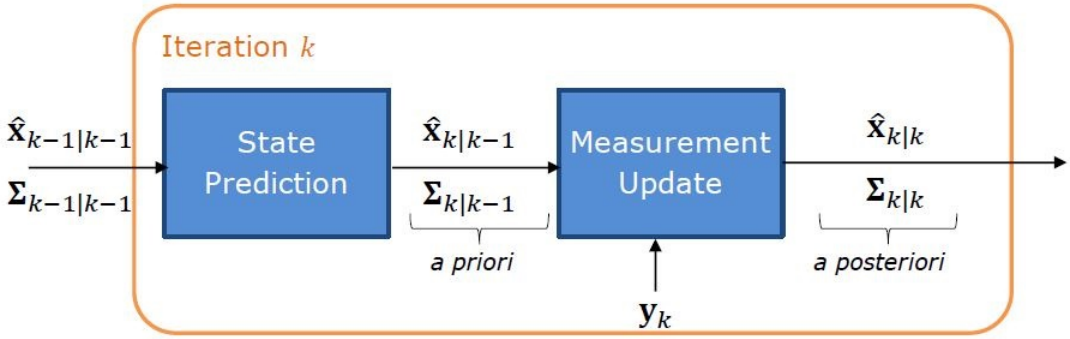


Figure 2.3: Kalman filter principles.

In case the transition model is based on non linear functions, the KF is called Extended Kalman Filter (EKF). In this case the state and measurements models (2.7) becomes:

$$\text{state transition model} \quad \mathbf{x}_k = \mathbf{f}_k(\mathbf{x}_{k-1}) + \mathbf{w}_{k-1} \tag{2.13}$$

$$\text{measurement model} \quad \mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k \tag{2.14}$$

This need to be linearised around the last estimated state $\hat{\mathbf{x}}_{k|k}$ so that a linear system is obtained and the previous procedure can be applied. The validity of he EKF solution depends from the uncertainty associated to the linearisation points and from the non-linearity. The EKF usually operates in a closed-loop mode which means that, every epoch the state vector is updated, it is used to correct the current state. Therefore the state vector prediction in equation 2.8 is no longer necessary and only the covariance propagation must be performed in prediction step. Details about KF and EKF can be found in [24], [25] and [26].

2.4 Outlier rejection

In the previous sections the measurements have been considered affected by noise, i.e. some accidental errors which follows a gaussian distribution. Unfortunately, when one perform numerous measurements is possible to find and observation that clearly differs from all the others. In this case is said that this observation is an outlier [27]. These ouliers can severely disturb the estimation procedure and so must be detected. When this difference is not so clear and affect an huge sample of measurements, it is necessary to use some robust outlier rejection method.

There are many types of robust algorithms that can be chose in function of the specific case study [28]. Among the many, the most interesting algorithm for this work is the RANdom SAmple Consensus (RANSAC) of Fischler and Bolles [29]. In general, the algorithm consists in robustly fitting a chosen model to a set of data that contains outliers. This is done in an iterative procedure of re-estimation and minimization of the chosen model starting from a random selection of minimal subset of data, sufficient to determine the model. Minimal subset means the number of sample N sufficiently high to ensure with a probability p that at least one of the random samples of s points is free from outliers. If the model is a line, the minimal subset consist in two points while if the model is a 2D projective transformation between two correspondent images, the minimal subset is represented by four correspondences. The reason is that the computational requirement to try every possible samples is often infeasible for large datasets. Usually a squared spatial distance of all the dataset with respect to this random model estimation is minimized and therefore, the best model from all the possible trial is selected in function of the number of outliers detected.

Given a set of data S containing outliers, estimate a variable that robustly fit a model. Procedure:

- Randomly select a sample of s data points from S and instantiate the model from this subset.
- Determine the set of data points S_i which are within a distance threshold t of the model. The set S_i is the consensus set of the sample and defines the inliers of S.
- If the size of S_i (the number of inliers) is greater than some threshold T, re-estimate the model using all the points in S_i and terminate.
- If the size of S_i is less than T, select a new subset and repeat the above.
- After N trials the largest consensus set S_i is selected, and the model is re-estimated using all the points in the subset S_i.

This algorithm is a fundamental procedure for any estimation problem which cope with a large quantity of outliers.

Chapter 3

METHODS FOR POSITIONING AND NAVIGATION

This Chapter gives an overview on basic principles and theoretical aspects of positioning and navigation estimation from different technologies and systems. In particular, the thesis combines different scientific disciplines connected to the navigation problem which are considered the prevalent in the field and the most suitable for the purpose of the work: the seamless navigation in challenging environments.

These are:

1. UWB positioning and navigation;
2. Image based positioning and navigation;
3. GNSS positioning with smartphone technology;
4. Inertial Navigation;
5. Other hybridization systems;

Each section of this chapter introduce the fundamental concepts, the methods and the algorithms of the related systems together with advantages and drawbacks in the localization task. The panorama is huge and includes disciplines like Computer Vision, Photogrammetry, Signal processing and Navigation. The motivation is that achieving robust positioning and navigation in challenging environments requires a large number of subsystem. In fact to overcome errors and noises, different sensors must be integrated to obtain a robust and reliable solution. Moreover, these systems needs fusion algorithms not only able to provide the optimal estimate for the single subsystems but also manage the relative interaction and the input/output priority. All this demonstrate the multidisciplinarity of the navigation problem and the expertise required to manage all the issues that any single system can inject in the overall solution.

3.1 Ultra-wide band systems

UWB is a radio technology for short-range, high-bandwidth communication, born for military purpose and used in distance estimation, localization and tracking. The use of this technology for positioning has increase a lot from 2002 when the United States FCC has regulated its use for communication, navigation and safety applications in mainstream markets [30]. Before 2002 the communication technologies providers started to be interested in commercial applications of this new technology due to his spectrum characteristics. In fact, the band allocated to this radiofrequency is 7.5 GHz wide, enormously greater than any other license given to the primary allocation groups (UTMS and IEEE 802.11 a / b / g). Furthermore, the use of this band is totally free. This regulation has therefore served to respond to the concerns and needs of hostile groups worried about the impact of this new market in their sectors. The details on this regulation are reported in figure 3.1 where the main point is the limitation of the power spectral density emission at the very low value of -41 dBm/MHz, in order to avoid interferences with other RF transmissions.

The definition and standardization of the UWB transmission systems is reported by the FCC and ITU-R which define it as "... a transmission from an antenna for which the emitted signal bandwidth exceeds the lesser of 500 MHz or 20% of the center frequency." In 2002, the 3.1 -10.6 and 22–29 GHz bands were opened to UWB by FCC, and the power spectral density emission for UWB transmitters was limited within −41.3 dBm/MHz (respectively 6.0 GHz 8.5 GHz in accordance to the European Communications Committee (ECC)).

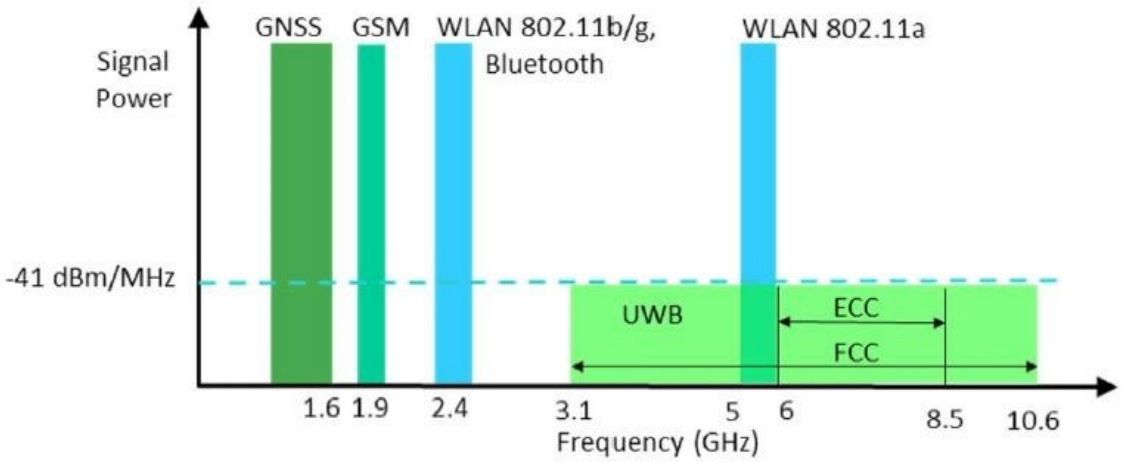


Figure 3.1: UWB spectrum regulation by FCC and ECC

In very general, an UWB transmission is a radio wave signal spread in the environment from a transmitter and received by an antenna receiver. These signals, defined as impulse radio Ultra-wideband (UWB-IR), are time-modulated ultra-wideband communications based on discontinues emission of very short pulses rather than continuous RF wave. In fact, combining multiple signals with a slightly different frequency, its possible to create a pulse with high defined timing, i.e., the peak of the pulse. Sinusoidal signal are narrow in frequency and wide over time while a pulse is narrow in time with a wide bandwidth in the frequency domain (Figure 3.2).

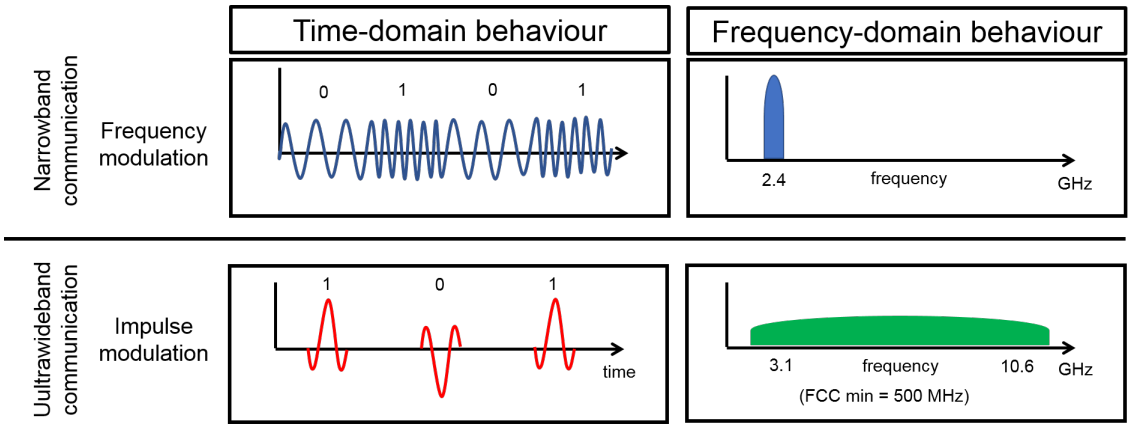


Figure 3.2: Narrowband vs Ultra-wide band communication behaviour. Time and frequency domain.

These characteristics gives to UWB-IR signal some important advantages in both the telecommunication and the positioning field. In telecommunication, the UWB-IR noiselike allows the signal to be shielded by interception and detection. Then due to these very short duration pulses, the UWB signal characteristic allows to mitigate multipath reflections. The huge bandwidth allocated permits to achieve high data rates communications. Moreover, due to the low-power spectral density assigned by law, UWB signals cause very little interference with existing narrow-band radio systems like WiFi and Bluetooth and can be emitted with very low power consumption. All these advantages are reflected in the ease implementation and in the low cost of these systems. It will shows later in this chapter, that some of these UWB signal characteristics will be also useful in the development of positioning and navigation solutions.

3.1.1 Principles of UWB positioning

Since this recent regulation, both the scientific community and the industrial sector have promoted the research and development of new systems that exploit

the physical characteristics of this technology. Indeed, the very short time-domain pulses of UWB systems make them ideal candidates for positioning application. As it was shown, the duration of a pulse is inversely proportional to the bandwidth of the transmitted signal. In UWB positioning the antenna receiver try to estimate the ToA of a pulse that, being very short, allows very little uncertainty on the time measurements. Knowing signal reception time with little uncertainty, then it is possible to estimate accurately the value of the distance travelled by the pulse between the two antennas simply multiplying the time of flight measurement by the speed of radio waves (speed of light c = 299792458 m/s). By combining this distance estimates at multiple receivers, it is possible to use simple triangulation techniques to estimate the position of the source. In general, knowing the position of some fixed tag and measuring a certain distance, it is possible to state that the unknown point will be in a circle of that radius around the anchor. The range between sensor i and the antenna receiver is given by:

$$\sqrt{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2} = c\,(t_i - t_0) \qquad i = 1,2,..,n \tag{3.1}$$

Where $(x,\ y,\ z)$ and $(x_i,\ y_i,\ z_i)$ are the coordinates of the tag and the sensor respectively, c is the speed of light, t_i is the signal ToA at sensor i, and t_0 is the unknown transmit time at the tag/device. Making distance measurements with 3 anchors, the position is uniquely determined by the intersection of the three circles (Figure 3.3). This is a general approach but many techniques of positioning based on signal time measurements are used in literature (Signal strength, Angle of Arrival (AoA), ToA, Time Difference of Arrival (TDoA)). For practical applications, the tag position should be estimate with a robust non-linear optimization algorithm.

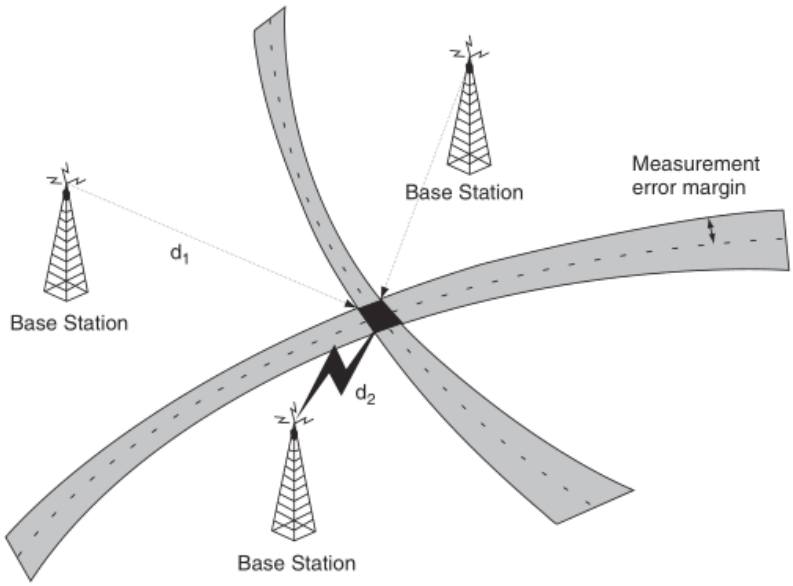


Figure 3.3: General triangulation example.

The use of large frequency bandwidth (>500 MHz) imposed by the regulation permit high resolution in time and consequently in range. In fact, in a single nanosecond, a wave travels almost 30 cm. Determinate the peak of a wide pulse accurately, is very hard and moreover the reflections coming from the signal scattered onto object (wall, ceilings, closets, desk, etc.) are also captured by the receiver and may overlap with line of sight pulse. With pulse of 4ns wide, ($\Delta f = 20MHz$) any object within 1.2 m of the receiver or the transmitter will cause an overlapping pulse. Because of this, ranging with Wi-Fi using Time of Flight (ToF) is not suitable for indoor applications. In the Ultra-Wideband systems, the duration of the pulse is in the order of nanoseconds or even less, that means a bandwidth of more than 500MHz and the capability to distinguish reflections of the signal.

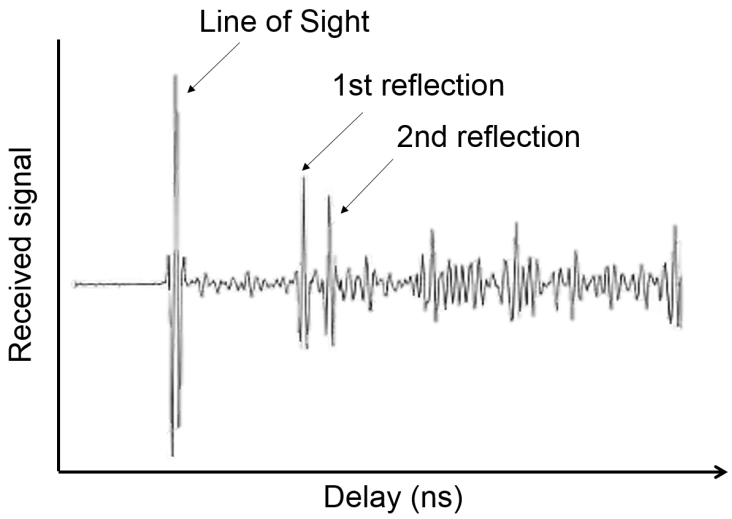


Figure 3.4: Waveform at the receiver of transmitted pulse. Analysing the peaks it is possible to discretize the reflections.

The achievable range resolution, in the case of UWB can be approximated with:

$$rr \approx \frac{c}{2b} \tag{3.2}$$

Where c is the speed of the wave front and b the bandwidth. E.g. for 500 MHz band and propagation in free space (assuming speed of light) it is $rr \approx 0.5c_0/500MHz = 299mm$. Whit any UWB systems, therefore, it is potentially possible to achieve sub-metre accuracy in poisoning. In the UWB systems, the high resolution time is measured with different techniques like ToA, Two Way Ranging (TWR), TDoA. All these techniques relies on time measurements and

has different advantages and disadvantages. The technique used in the systems analysed is the TWR.

The TWR methods can be described as follows. A transceiver (from now on Anchor) transmits a UWB radio message at a certain time (recorded timestamp t_1) Another transceiver (form now on Tag), usually moving in the environment, receives the message and transmits a response back to the anchor after a particular fixed time delay t_{replay}. The anchor then receives the tag response at a certain time and records it (recorded timestamp t_2). Thanks to the two timestamps t_1 and t_2, the anchor is able to calculate the so called round trip time $t_{roundtrip}$. Theoretically, subtracting to this time, the time delay of the tag response and dividing this by 2 (due to the $r_{roundtrip}$) it is possible to determinate the ToF of the message. Being the radio wave a very short pulse, the time accuracy will be very high. Then, assuming the speed of a radio wave through the air equals to the speed of light c, then the distance between the anchor and the tag can be calculated. It is important to note that the only reference time for the TWR computation is the one of the anchors timestamps which means that the non-synchronizations of the transceivers inner clocks is not a problem (Figure 3.5).

$$t_{roudtrip} = t_2 - t_1 \tag{3.3}$$

$$TOF = \frac{t_{roundtrip} - t_{replay}}{2} \tag{3.4}$$

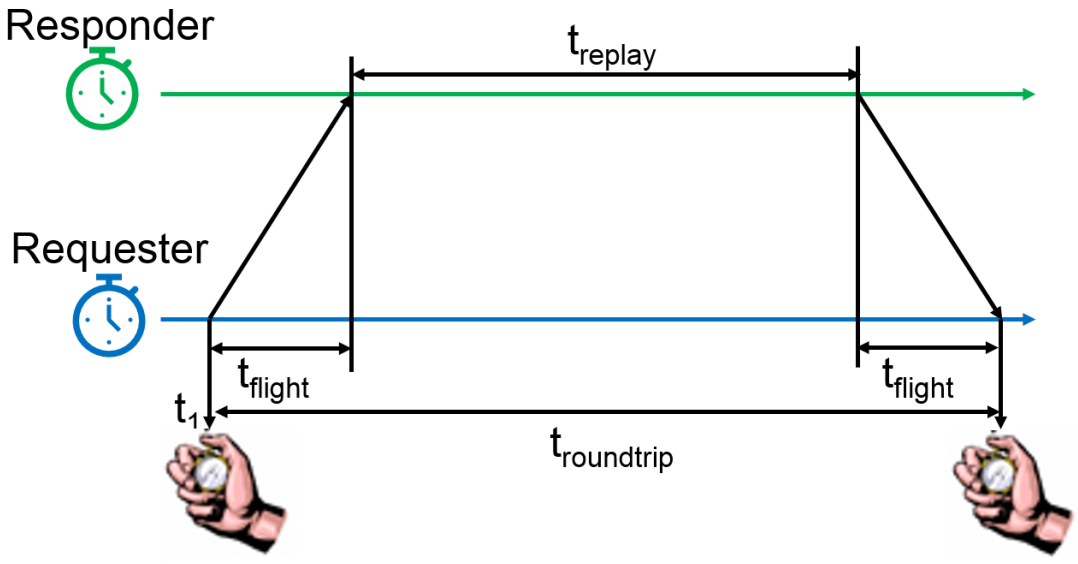


Figure 3.5: Two-Way Ranging description.

3.1.2 Source of UWB positioning errors

Propagation of UWB signal suffer from environmental condition while physical implementation of the sensor can affect the measurements. In general, the sources of positioning uncertainty in UWB systems are:

- oscillator accuracy and drift;
- walls and obstructions that increase path length;
- received signal strength;
- accuracy of delay estimation technique;
- false readings from interference and multipath;
- saturation of the communications.

For very high accuracy in positioning these sources of errors should be avoided, in particular in Non Line of Sight (NLOS) conditions. In these situations, the estimation procedure must be supported by other techniques like weighting error region based on confidence levels or logarithmic likelihood measure based on signal to noise ratio.

3.2 Image-based systems

The interest in digital images as primary data for positioning and navigation purpose raise from the possibility to extract not only spatial information simulating human perception, but also the geometry of the three-dimensional real word and the spatial relation with it. With such information is possible to produce accurate topographic maps, to reconstruct the reality in three dimensional models (from an architectural scale till the industrial metrology), to define the pose of sensors that acquires such images and calibrate them, to recognize and track objects in the scene and much more.

In the following sections, the principles and the mathematical techniques useful to understand the proposed algorithms for image-based positioning are presented. These are basic principles useful for the algorithm implementation in further chapters. In Chapter 4 and 5 will be presented two case studies of image-based positioning and navigation based on two different approaches.

The first is a problem of camera pose estimation exploiting photogrammetric principles while the second is a visual odometry problem in which computer vision methods were used. The reason of this dual approach is inherent to the academic background of the author which has been influenced from the Geomatics research group of Politecnico di Torino (for the photogrammetric approach) and from the

Sensor and Indoor Navigation (SINA) research group of Finnish Geospatial Research Institute (FGI) (for the computer vision approach).

Although both fields have considerable overlap, is possible to define different goals between them, in particular computer vision has evolved under the central theme of achieving human level capability in the extraction of information from image data [31] while photogrammetry found his main core in the accuracy. Moreover, while computer vision express the basic camera model as a linear mapping of homogeneous coordinates, in photogrammetry this is usually expressed in terms of collinearity equations.

It is not topic of this work going in deep on this comparison, what is important to state is that whit the information technology revolution, complex and heavy computational procedure have become more fast and accurate, allowing real-time applications. Moreover, uncalibrated low-cost sensors has been introduced in a field that before was restricted to tools of higher value.

To understand better this duality of the problem and how it will be correlated with the present work, let's consider two different situations:

- A camera moves while acquiring a static scene (singular view);
- Two pictures of the same scene taken at different time (stereo view).

These two scenarios represent the same geometric problem but in this Thesis represent also the two different procedures made to perform positioning and navigation with images. Figure 3.6 shows better this concept: in the so called "Computer Vision approach", a monocular camera moves while acquiring a static scene. Knowing simply some initial information is possible to perform a Visual Odometry procedure to estimate the navigation path of the camera. This means updating the pose of the camera iteratively from the previous position. Moving on the "Photogrammetric approach", a camera acquire an image of an object which is already documented in a complete database of geolocated images. From the collinearity equations and knowing the distance of the database images from the object is possible to estimate the position of the new camera.

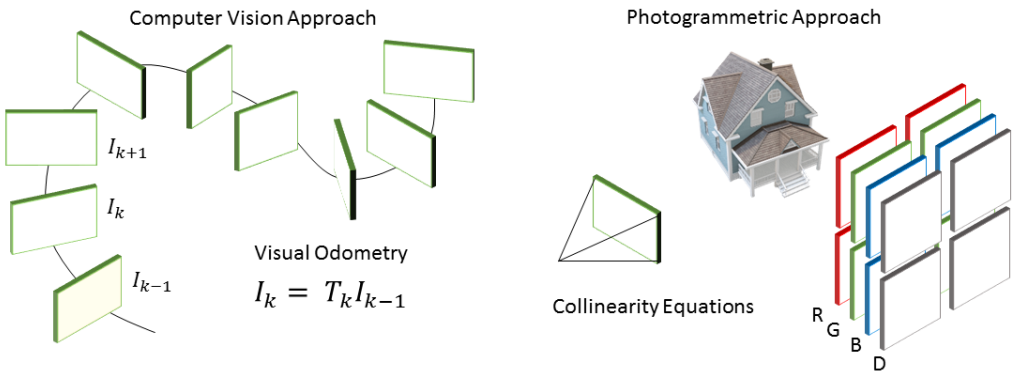


Figure 3.6: Schema of computer vision vs photogrammetric approach.

3.2.1 The computer vision approach

This Section explain the principles of computer vision and how this research field is related to the navigation problem. In very general term computer vision is the study of visual data acquired by mechanical and digital sensors that simulate the human vision. This discipline has enormously increased his importance in the last years as a result of the large number of camera sensors spread all over the world. Nowadays everyone carries around smartphones with one, two or maybe even three cameras mounted on it. This means that billions of sensors are acquiring continuously a large massive amount of visual data that need to be analysed and processed for different purpose. Computer vision is an interdisciplinary field that operates on different areas of science, engineering and technology. It touches areas like physics, biology, psychology, computer science, mathematics and more. (Figure 3.7)

The first step to automatically understand the content of visual data is to understand optics and image formation. The history of computer vision born exactly with the tentative to reproduce mechanically the human vision, in particular the eyes image formation, through a camera obscura based on pinhole camera theories. The second fundamental step in computer vision was understanding the animals visual perception mechanism. Hubel and Wiesel in late 50s discovered that the cats brain can recognize the complex visual world from oriented edges and simple structure observed in the real world [32]. This gave the intuition to simplify the real world in simple geometric shape and to relate it with complex interpretation and reconstruction of the three-dimensional reality. From the 60s on, researcher start to investigate on algorithm able to perform object detection and recognition in digital images and use it as information for modelling the geometry of multiple views cameras [33] [34]. The third challenging step in computer vision was to estimate the motion from the camera images as an animal or an human do with their

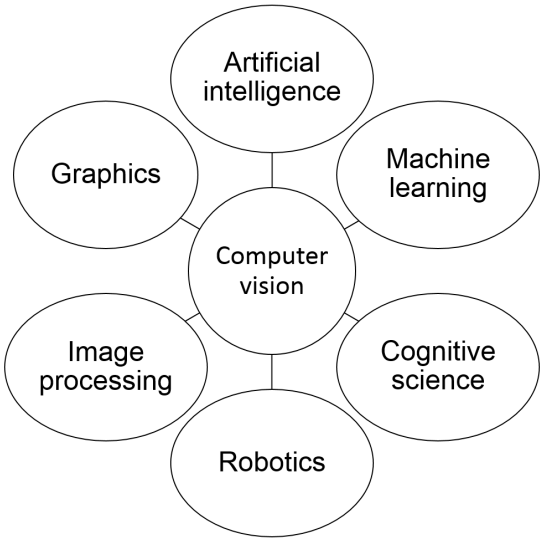


Figure 3.7: The computer vision interdisciplinarity.

eyes. Thank to the visual perception of the brain is easy for humans to understand the perspectives, to retrieve missing information from occluded parts in the scene and to estimate distances to the objects. The earliest form of motion estimation in computer vision is "optical flow", a concept arises from many studies performed on bees in a wind tunnel [35]. After that, the robotic community has investigated on navigation systems based on Visual Odometry, the field of investigation that try to retrieve the trajectory of a camera or of a stereo camera rig using the visual information acquired in the images. Deduction of motion information and position from images is the main goal of this chapter. In particular, estimating camera trajectory in 3D and the 3D position of the feature point based on the apparent motion of 2D correspondences is a problem in numerous filed of investigation like:

- camera tracking;
- structure from Motion;
- photogrammetric mapping;
- simultaneous Localization and Mapping (SLAM).

3.2.1.1 Pinhole camera model

An image sensor in a camera is an array of light sensitive detectors that converts 3D world information in 2D pixels on an image plane. This image i.e. the photographs, contains digital information like colours and intensity value inside each pixel and can be used as input data in computer vision algorithm. The relation

between the 3D scene in the real world and the recorded image in the 2D image plane is called camera model. Each three-dimensional point in the field of view of the camera is mapped into a two-dimensional feature using a projective transformation. The mathematical model that allow to describe this projection is composed by a sequence of transformation between different coordinate frames and projective space expressed by matrix equations. The following schema is an overview on all these equations seen both as forward projection from 3D scene to 2D camera features then as backward projection from 2D pixel coordinates to 3D world points. This second case is what CV is usually interested in as allow to recover 3D scene structure simply from image sequences.

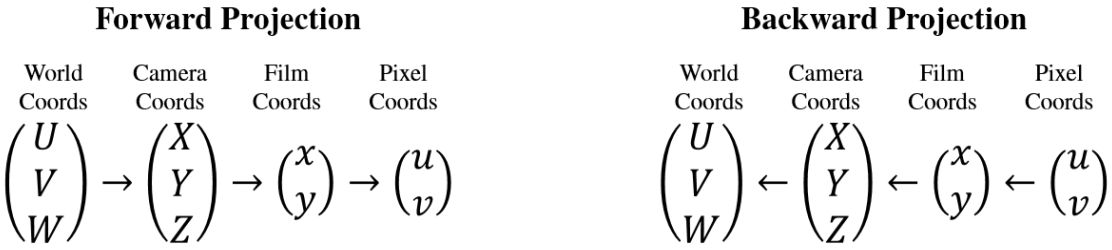


Figure 3.8: Camera projections

The simplest camera model is the pinhole camera model (Figure 3.9) which correspond to the central projection where a point in space $\mathbf{X} = (X, Y, Z)^T \in \mathbb{R}^3$ is projected onto an image plane along a ray that connect this point to the camera centre of projection **C**. The 2D point in the image plane is called image point $\mathbf{x} = (x, y)^T \in \mathbb{R}^2$. The general name of the pinhole camera model is the "3D to 2D perspective projection".

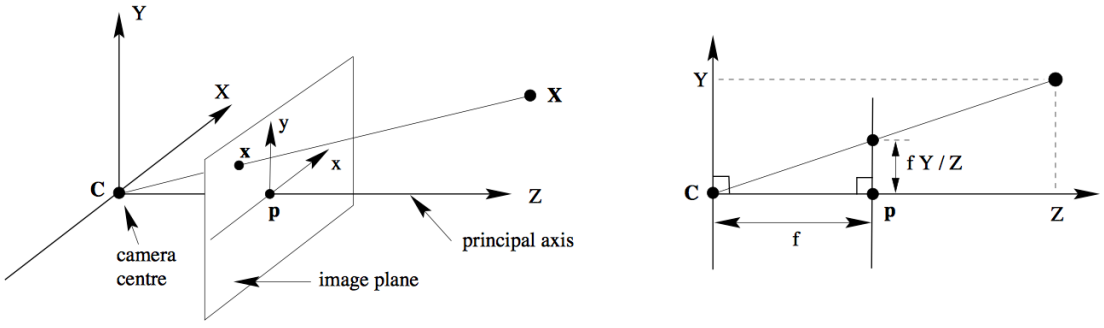


Figure 3.9: Pinhole camera model [34].

If **C** is the origin of an Euclidean coordinate system with the Z axis in the direction of the camera view, the point where Z axis meets the image plane (at distance f from **C**) is called principal point **pp**. f is the physical focal length,

the distance between **C** and the image plane. In this Euclidean geometry called camera frame, the 3D scene point $\mathbf{X}_{cam} = (X, Y, Z)$ correspond to a 2D image point $\mathbf{x}_{cam} = (x, y)$ by similarity rules of the triangles composing the geometric model of the pinhole cameras (central projection, dimension of an object in an image is inversely proportional with the depth or distance Z):

$$x = f\frac{X}{Z} \qquad y = f\frac{Y}{Z}$$
$$(X, Y, Z)^T \longmapsto (fX/Z, fY/Z)^T \tag{3.5}$$

In order to represent these relations as matrix equations is necessary to introduce the homogeneous coordinate that allow us to define every 2D points $(x, y)^T \in \mathbb{R}^2$ as a 2D triple point $(kx, ky, k)^T$ in $\mathbb{P}^2$ adding a fictional third coordinate ($k = 1$). Reversely, for a given triple $(\widetilde{x}, \widetilde{y}, \widetilde{z}) \in \mathbb{P}^2$ in homogeneous coordinate we can recover his Cartesian form $(x, y) \in \mathbb{R}^2$ as $x = \widetilde{x}/\widetilde{z}$ and $y = \widetilde{y}/\widetilde{z}$. The same is valid for 3D scene point $(X, Y, Z)^T \in \mathbb{R}^3$ which can be expressed as $(X, Y, Z,1)^T \in \mathbb{P}^3$. In this context, the perspective transformation 3.5 can be expressed:

$$\begin{pmatrix} \widetilde{x} \\ \widetilde{y} \\ \widetilde{z} \end{pmatrix} = \begin{bmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}; \quad \begin{pmatrix} \widetilde{x} \\ \widetilde{y} \\ \widetilde{z} \end{pmatrix} = \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} \tag{3.6}$$

that express the 3D to 2D forward projection

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \in \mathbb{P}^3 \longrightarrow \begin{pmatrix} fX \\ fY \\ Z \end{pmatrix} \in \mathbb{P}^2$$

The matrix in this expression is a (3×4) and may be written as $M_{3\times4} = diag\,(f, f,1)\,[I \mid \mathbf{0}]$ and the expression 3.6 has the concise form:

$$\mathbf{x} = diag\,(f, f,1)\,[I \mid \mathbf{0}]\mathbf{X}_{cam} \tag{3.7}$$

Then is possible to recover back the cartesian coordinates (x, y) in $\mathbb{R}^2$ by dividing the point by his last value:

$$x = f\frac{X}{Z} \qquad y = f\frac{Y}{Z} \in \mathbb{R}^2$$

Until now it has been assumed that the origin of the image plane coordinates coincides with the principal point but this in practice may not be. In real case, what it possible to measure on images are pixel values which have their own origin

of coordinate system, usually in the top-left corner of the sensor. The actual pixel values are:

$$x_{pix} = \frac{x}{d_x} + u_0 \equiv u \quad \text{and} \quad y_{pix} = \frac{y}{d_y} + v_0 \equiv v$$

With d_x and d_y pixel size in world unit, $O = (u_0, v_0)$ shift of the top-left origin with respect to the camera center and u and v pixel coordinate of the point in the image frame. In matrix form and homogeneous coordinate equation 3.6 becomes:

$$\begin{pmatrix} \widetilde{x}_{pix} \\ \widetilde{y}_{pix} \\ \widetilde{z}_{pix} \end{pmatrix} \equiv \begin{pmatrix} \widetilde{u} \\ \widetilde{v} \\ \widetilde{w} \end{pmatrix} = \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} \widetilde{x} \\ \widetilde{y} \\ \widetilde{z} \end{pmatrix}; \quad \begin{pmatrix} \widetilde{u} \\ \widetilde{v} \\ \widetilde{w} \end{pmatrix} = \begin{pmatrix} \frac{1}{d_x}\widetilde{x} + u_0 \\ \frac{1}{d_y}\widetilde{y} + v_0 \\ \widetilde{z} \end{pmatrix} \tag{3.8}$$

The matrix in this expression is a (3×4) representing an affine transformation. Summarizing, between a point expressed in the image plane expressed in camera coordinate and the same point expressed in pixel coordinate there is a 2D-to-2D Affine transformation, while a 3D point in homogeneous coordinate is mapped into a 2D point in homogeneous coordinate, (both expressed in camera frame) with a 3D-to-2D Projective transformation. Joining the two transformation is possible to express the projective transformation that maps scene points in camera frame to image point expressed in pixel coordinates:

$$\begin{pmatrix} \widetilde{x}_{pix} \\ \widetilde{y}_{pix} \\ \widetilde{z}_{pix} \end{pmatrix} \equiv \begin{pmatrix} \widetilde{u} \\ \widetilde{v} \\ \widetilde{w} \end{pmatrix} = \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{3.9}$$

$$\begin{pmatrix} \widetilde{u} \\ \widetilde{v} \\ \widetilde{w} \end{pmatrix} = \begin{bmatrix} \frac{f}{d_x} & & u_0 & 0 \\ & \frac{f}{d_y} & v_0 & 0 \\ & & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{d_x} fX + u_0 \\ \frac{1}{d_y} fY + v_0 \\ Z \end{pmatrix} \tag{3.10}$$

The concise form for this expression is:

$$\mathbf{x}_{pix} = M_{3\times3} \cdot M_{3\times4} \mathbf{X}_{cam} \tag{3.11}$$

Then is possible to recover back again the Cartesian coordinates $\mathbf{x}_{pix} = (u, v)^T$ in $\mathbb{R}^2$ by dividing the point by his last value:

$$\Longrightarrow \qquad u = \frac{\widetilde{u}}{\widetilde{w}} = \frac{1}{d_x} f \frac{X}{Z} + \frac{u_0}{Z} \qquad v = \frac{\widetilde{v}}{\widetilde{w}} = \frac{1}{d_y} f \frac{Y}{Z} + \frac{v_0}{Z} \qquad \in \mathbb{R}^2$$

Is possible to observe that all the parameters related to the inner characteristic of the camera are expressed inside this two matrix. In fact the projective transformation express the zooming effect of the lenses in the image formation while

the affine transformation take into account the main mechanic proprieties of the camera sensor. Notice that from 3.2.1 it is possible to define a matrix that not depend at all from the scene observed by the camera:

$$K = \begin{bmatrix} f_x & & u_0 \\ & f_y & v_0 \\ & & 1 \end{bmatrix} \tag{3.12}$$

This (3×3) matrix is called camera calibration matrix and is composed by the internal parameters of the camera.

Then the **pinhole camera model** can be expressed as:

$$\mathbf{x} = K\,[I \;\mid\; \mathbf{0}]\,\mathbf{X}_{cam} \tag{3.13}$$

With $\mathbf{x}$ and $\mathbf{X}_{cam}$ both expressed in homogeneous coordinates. The Euclidean coordinate system centered in the camera centre with the principal axis of the camera pointing down the Z-axis is called camera coordinate frame and in the previous equations the 3D scene points were expressed in this coordinate system ($\mathbf{X}_{cam}$). In general, the point in the object space are expressed in a different Euclidean coordinate frame, the world coordinate frame. As is known, two frames are related via a rigid transformation composed by rotation, translation and scaling.

$$X_{cam} = RX_{world} + t \tag{3.14}$$

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}_{cam} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}_{world} \tag{3.15}$$

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}_{world} = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}^{-1} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}_{cam} \tag{3.16}$$

where R is a 3×3 rotation matrix and t is a 3×1 column vector. So from Equation 3.13, the relation between the scene points in the world frame and the image points in the camera frame can be written now:

$$\mathbf{x} = K\,[R \;\mid\; \mathbf{t}]\,\mathbf{X} \tag{3.17}$$

$$\begin{pmatrix} \widetilde{u} \\ \widetilde{v} \\ \widetilde{w} \end{pmatrix} = \begin{bmatrix} \frac{1}{d_x} & 0 & u_0 \\ 0 & \frac{1}{d_y} & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{3.18}$$

Where $\mathbf{X}$ now is expressed in the world coordinate frame. The parameters of the camera calibration matrix are called internal camera parameters and are independent form the real word, while the parameters of the roto-translation matrix are called external parameters or exterior orientation and relate the camera center position and orientation to a world coordinate system. The combination of these two transformation is expressed in the (3×4) camera matrix P:

$$P = K\,[R \;\mid\; \mathbf{t}] \tag{3.19}$$

from which the pinhole camera model is:

$$\mathbf{x} = P\mathbf{X} \tag{3.20}$$

Remember that all the equation matrix expressed above are up to scale equivalence and in order to make the theoretical description more light the following equivalence has been used: $\widetilde{x} \rightarrow x = (\widetilde{u}, \widetilde{v}, \widetilde{w})^T$.

3.2.1.2 Lens Distortion

In real cameras, the pinhole model is not respected as the lenses and mechanic characteristic of the camera sensor introduces some deviations in this model. In particular, the wide angles of the lenses result in a radial distortion in the images which cause straight lines in the scene to show up as curved lines in the image (Figure 3.10).

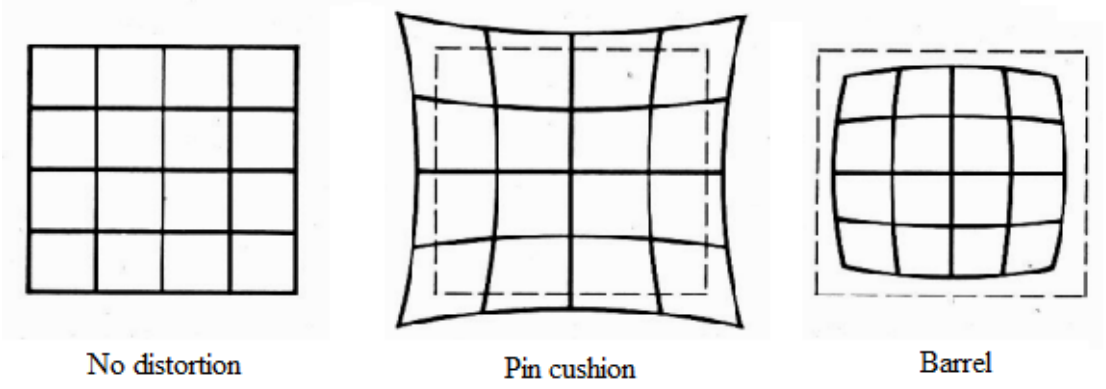


Figure 3.10: Distortion due to optics.

The amount of this distortion on the radial distance from the centre of the image is expressed as follow:

$$\begin{bmatrix} \widetilde{x}_{dist} \\ \widetilde{y}_{dist} \end{bmatrix} = \left(1 + k_1 \cdot r^2 + k_2 \cdot r^4 + k_3 \cdot r^6\right) \begin{bmatrix} \widetilde{x}_{und} \\ \widetilde{y}_{und} \end{bmatrix} \tag{3.21}$$

With k_i distortion value function of the specific camera and $r = \widetilde{x}_{und}^2 + \widetilde{y}_{und}$. Moreover the camera can also be affected by tangential distortion caused by physical elements in a lens not being perfectly aligned:

$$\widetilde{x}_{und} = x_{dist} + [2p_1 y_{dist} + p_2 \left(r^2 + 2{x_{dist}}^2\right)] \tag{3.22}$$

$$\widetilde{y}_{dist} = y_{dist} + [p_1 \left(r^2 + 2{y_{dist}}^2\right) + 2p_2 x_{dist}] \tag{3.23}$$

With p_1 and p_2 tangential distortion coefficients of the lens. To determinate the distortion parameters $(k_1, k_2, k_3, p_1, p_2)$ useful to correct the lens distortions, one may use several procedures which exploit images of a known calibration object such as a chequerboard or a calibration field.

3.2.1.3 Camera Calibration

Camera calibration procedure is the process that allows to estimate intrinsic and extrinsic camera parameters from a set of 3D to 2D-points correspondences (resectioning), including the distortion coefficients, using non-linear least-square minimization. In section 3.2.1 the intrinsic parameters of an ideal camera were saved in the K matrix. For real CCD cameras, there is the possibility to have non-square pixels which introduce a unequal scale factor $(\alpha_x \ \alpha_y)$ in the calibration matrix. Moreover, in some cameras could be presented a skew coefficient between the image plane axis (S). Taking into account also these parameters, the camera calibration matrix is composed by focal length (f_x, f_y), principal point (u, v), skew coefficient (S) and aspect ratio $(\alpha_x \ \alpha_y)$. The camera external parameters are six parameters, three for the rotation and three for the translation. To calibrate a camera there are different state-of-art procedure that can be classified in linear and non-linear approaches. Starting from linear problem:

From equation 3.20, if a set of points in the world reference frame is known (for example from a calibration rig or a checkerboard) and the same point can be detected in the image, it is possible to set up a linear system of equations from n correspondences $X_i \leftrightarrow x_i$:

$$\begin{bmatrix} \mathbf{0}^T & -\omega_i \mathbf{X}_i^T & y_i \mathbf{X}_i^T \\ \omega_i \mathbf{X}_i^T & \mathbf{0}^T & -x_i \mathbf{X}_i^T \end{bmatrix} \begin{pmatrix} \mathbf{p}_1 \\ \mathbf{p}_2 \\ \mathbf{p}_3 \end{pmatrix} = 0 \tag{3.24}$$

$$\mathbf{Ap} = \mathbf{0} \tag{3.25}$$

where each $\mathbf{p}_i^T$ is a 4-vector for the i-th row of $\mathbf{P}$ This system has $2n$ equation from a set of n correspondences between object points and image points and 12 unknowns as P is a (3×4) matrix. When the homogeneous linear system is overdetermined $(2n > 11)$ and there is the presence of noise in the point coordinates, there is no exact solution to the equations $\mathbf{Ap} = \mathbf{0}$ but an approximate solution that minimize

a suitable cost function (algebraic or geometric distance). This is the Direct Linear Transformation procedure and is valid also to compute the projective transformation between 2D-to-2D image point correspondences. One possible solution of this minimization problem can be stated as follow:

given a matrix A with at least as many rows as columns, find x that minimize $||Ax||$ subject to $||x = 1||$, the approximate solution x is the last column of V, where $A = UDV^T$ is the singular value decomposition of A. The obtained solution is P, the camera projection matrix that best fit the points correspondences. Therefore is possible to find the camera center,the orientation of the camera and the internal parameters of the camera.

Having **P** like in 3.19, it is possible to write:

$$\begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} = \begin{bmatrix} f_x & & & u_0 \\ & f_y & & v_0 \\ & & 1 & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \tag{3.26}$$

And linearly solving the equations to find all the unknown. This is the full Camera calibration process which estimates the values of the intrinsic parameters, the extrinsic parameters, and the distortion coefficients from P.

A more common procedure to estimate intrinsic and extrinsic parameter from a camera matrix is based on the Plane Based Calibration. It is based on the use of a calibrated checkboard as scene for the multiple acquisition of a single camera. (Figure 3.11) where the world frame is on the plane such that XY-plane correspond to a plane π in the scene.

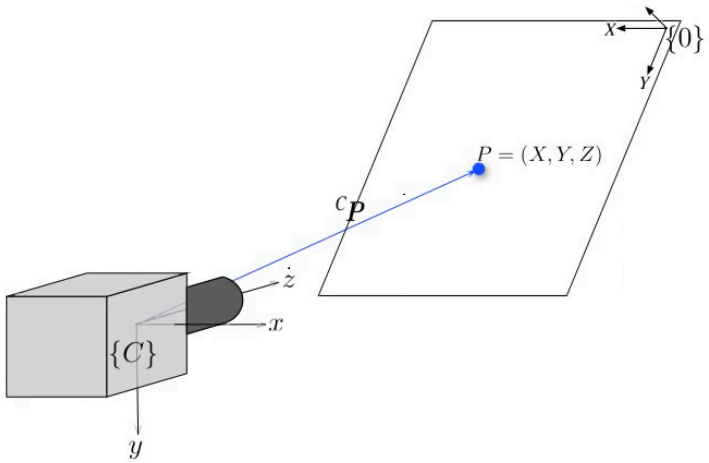


Figure 3.11: Plane induced homography.

When all the object 3D points lie on a plane ($Z = 0$) and the camera centre don't correspond to the scene plane, the mapping of this points on the image plane

can be expressed as:

$$\mathbf{x} = P\mathbf{X} = \begin{pmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ 0 \\ 1 \end{pmatrix} = [\mathbf{p}_1\ \mathbf{p}_2\ \mathbf{p}_4] \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \tag{3.27}$$

This projective camera matrix is an planar homography transformation (a plane to plane projective transformation) represented by H^p a 3×3 matrix. For a calibrated camera of the type in eq. 3.12, the homography between a world plane at $Z = 0$ and the image is:

$$H = K[\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{t}] \tag{3.28}$$

Where $\mathbf{r}_i$ are the column of R. As we have an homography projective transformation between 3D coordinate points and 2D image point for every camera position, and as the camera internal parameter are the same for each image acquisition its valid the relation:

$$H = \lambda K[\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{t}] \tag{3.29}$$

Making the inverse:

$$[\mathbf{r}_1, \mathbf{r}_2, \mathbf{t}] = \frac{1}{\lambda} K^{-1}[\mathbf{h}_1\ \mathbf{h}_2\ \mathbf{h}_3] \tag{3.30}$$

As $\mathbf{r}_1$ and $\mathbf{r}_2$ are column vectors of a rotation matrix, the norm $\| \mathbf{r}_1 \| = \| \mathbf{r}_2 \| = 1$ and $(\mathbf{r}_1)^T \cdot \mathbf{r}_2 = 0$. Using these as constrain it is possible to estimate R and $\mathbf{t}$ as follow:

$$\mathbf{r}_1 = \lambda K^{-1}\mathbf{h}_1 \tag{3.31}$$

$$\mathbf{r}_2 = \lambda K^{-1}\mathbf{h}_2 \tag{3.32}$$

$$\mathbf{r}_3 = \mathbf{r}_1 \times \mathbf{r}_2 \tag{3.33}$$

$$\mathbf{t} = \lambda K^{-1}\mathbf{h}_3 \tag{3.34}$$

Using this as initial parameter, at the end of a refinement procedure a calibrated camera is obtained. The refinement procedure consist in compute distortion coefficient and optimize initialization parameters using non-linear least square minimization:

$$\widehat{\mathbf{p}} = \underset{p}{argmin} \mid \mathbf{x} - f(\mathbf{p}) \mid$$

where $\mathbf{p}$ are the camera parameters and the function f project the known planar 3D points $\mathbf{X}$ onto 2D image plane with argument $\mathbf{p}$. The iterative method which minimize the chosen cost function is the Sparse Levenberg-Marquardt algorithm [36] which update the parameters starting from the initial estimate $\widehat{\mathbf{p}_0}$.

This estimation problem has many feature in common with two other important estimation procedure that will be described later: the 2D homography estimation

and the Fundamental matrix computation. If another image acquiring the same scene plane is considered, then it is possible to affirm that point correspondences x_i and x_i' in the two images are also related by a homography H. This special case is fundamental for the navigation solution proposed and will be presented in section 3.2.1.

3.2.1.4 Two-View Geometry background

The camera image acquisition can be considered to a first approximation, as a central projection where a point in the scene is mapped to the correspondent point in the image plane along a ray that connect the object point to the centre of projection (camera centre). If no information about the three-dimensional real world is given, it is possible to reverse the problem and say that the 3D point in the scene can lie everywhere in the straight line (or projecting ray) that starts from the camera centre and pass through the image point. If the same scene is framed by another image, the projection of the straight line generated by the first image on the second image represent the place where the object point can lie. This mean that for an image point x in the first view, the search of the correspondent x' point in a second image viewing the same 3D-space, can be constrained to a single line called epipolar line. This constrain increase the accuracy in the identification of image correspondences and decrease the computational cost of the process.

Suppose now a 3D-space point $\mathbf{X}$ viewed by two camera and projected to a 2D point $\mathbf{x}$ in the first view, and $\mathbf{x}'$ in the second. These three points are coplanar, and all lie on the same plane. The intersection between this plane and the two-image plane correspond exactly with the previous epipolar line. For each image point correspondences there is a plane defined by a line between the two-camera centre C and C' (baseline) and the object point $\mathbf{X}$.

If one camera is in the view of the other camera, the baseline intersects the image plane in a point called epipole. All these geometric relations are called epipolar geometry. (Figure 3.12)

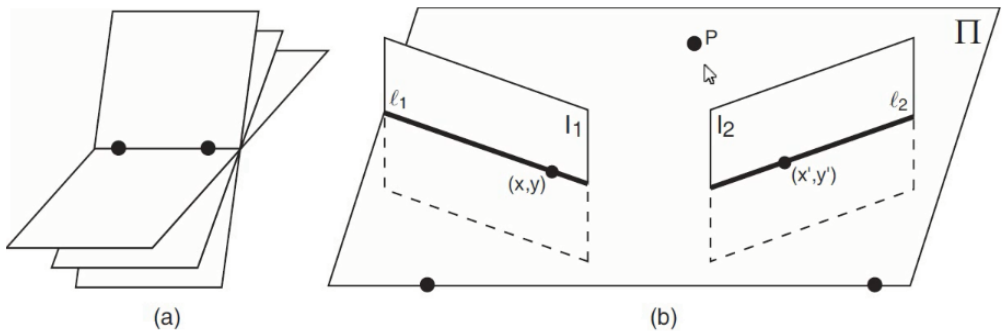


Figure 3.12: Epipolar geometry.

The epipolar geometry is represented by an algebraic relation called Fundamental Matrix expressed in homogeneous coordinates: *"For any set of correspondent points* $\{x_i \leftrightarrow x'_i\}$ *in the two images,* $x'^T\ Fx = 0$ *".*

$$\forall\ (x, y) \leftrightarrow (x', y') \qquad \mathbf{x}'^T\ F\mathbf{x} = 0 \tag{3.35}$$

Where the fundamental matrix is a (3×3) matrix of rank 2. Assuming fixed (x, y) in the first image:

$$\begin{bmatrix} x' & y' & 1 \end{bmatrix}^T F \begin{bmatrix} x & y & 1 \end{bmatrix} = 0 \qquad \begin{bmatrix} x' & y' & 1 \end{bmatrix}^T F \begin{bmatrix} a & b & c \end{bmatrix} = 0 \tag{3.36}$$

$ax' + by' + c = 0$ is the Epipolar line in the second image and vice versa $a'x + b'y + c' = 0$ is the Epipolar line in the first image. Therefore, the Fundamental matrix maps a 2D point in the image plane of the first view to a 2D line in the image plane of the second view:

$$\mathbf{x} \mapsto \mathbf{l}' \qquad \mathbf{l}' = F\mathbf{x} \tag{3.37}$$

Faugeras and Hartley in 1992 introduces the algebraic relation between the fundamental matrix and the ordered camera matrix pair $\{P, P'\}$. The fundamental concept to state is that:

"Given the ordered camera matrix pair $\{P, P'\}$ there is an unique (up to scale) associated matrix F while given a matrix F there is a family of camera matrices $\{PW, P'W\}$. Moreover: *"A non-zero matrix* F *is the fundamental matrix corresponding to a pair of camera matrices* P *and* P*' if and only if* $P'^T F P$ *is skew-symmetric."* From all this is always possible to bring one of the camera matrix to the form $P = [I \,|\, 0]$ through and SVD decomposition and so assume the second camera matrix as $P' = [B \,|\, b] = [R \,|\, t] = \left[[e']_\times F \,|\, e'\right]$.

The standard algorithm used to compute the fundamental matrix from point correspondences will be presented in next paragraph (8-point algorithm).

Certain case of motion or particularly favourable case allow to simplify the fundamental matrix computation. It has been already shown how to calibrate a camera and knowing camera parameter, normalized image coordinates may be used to compute the Fundamental Matrix. In this case the computed matrix is called Essential matrix E and it satisfy .

$$\forall\ (x, y) \leftrightarrow (x', y') \qquad \mathbf{x}'^T\ E\mathbf{x} = 0 \tag{3.38}$$

With:

$$E = K'^T F K. \tag{3.39}$$

3.2.1.5 Estimating the Fundamental Matrix

To estimate the fundamental matrix F having a pair of real images there are several procedures relying on linear and non-linear methods based on different cost functions or algebraic distances. In this paragraph will be described the general method based on a least-square solution of a set of linear equations (8-point algorithm). This estimation can be considered a first approximate solution that have to be introduced in an iterative procedure to consider noise on the measurements and outliers. To automatize and refine the estimation of the initial fundamental matrix an iterative method that rely on non-linear minimization and on robust outlier rejection should be used (RANSAC).

Given a set of image point correspondences $\{x_i \leftrightarrow x_i'\}$, determinate the fundamental matrix F such that ${x_i^i}^T \; Fx_i = 0$. Procedure:

- Normalization: Apply to the point correspondences a transformation to have image coordinate with zero mean and standard deviation equal to 1, $(x_i, x_i') \rightarrow (\hat{x}_i, \hat{x}_i')$;
- Linear equation: Each correspondence gives one constrain on F. With i, number of correspondences:

$$\begin{bmatrix} x_i' \\ y_i' \\ 1 \end{bmatrix}^T \begin{bmatrix} F_{11} & F_{12} & F_{13} \\ F_{21} & F_{22} & F_{23} \\ F_{31} & F_{32} & F_{33} \end{bmatrix} \begin{bmatrix} x_i \\ y_i \\ 1 \end{bmatrix} = 0 \tag{3.40}$$

$$\Rightarrow \begin{bmatrix} x_1'x_1 & x_1'y_1 & x_1' & y_1'x_1 & y_1' & x_1 & y_1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n'x_n & x_n'y_n & x_n' & y_n'x_n & y_n' & x_n & y_n & 1 \end{bmatrix} \begin{bmatrix} F_{11} \\ \vdots \\ F_{33} \end{bmatrix} = 0 \tag{3.41}$$

$$(n \times 9)(9 \times 1)$$

$$AF = 0$$

- SVD of A: compute the singular value decomposition of A as $A = UDV^T$ and set $\hat{F}$ as the smallest singular value of A (last column of V) reshaped in a (3×3) matrix. As the singular value of A could be not zero the SVD of $\hat{F}$ have to be computed in order to obtain the correct F;
- Denormalization: apply to the previous result a back-transformation in order to obtain the final fundamental matrix F.

A similar procedure, called five-point algorithm (Li and Hartely 2006) can be used to estimate the Essential matrix.

3.2.1.6 Degeneracy case: Planes and Homography

In chapter 3.2.1 it has been already introduced the concept of plane induced homography, when the world scene acquired by a camera views is a planar surface (Eq. 3.28). Let's suppose now that another camera with the same parameters acquire the same planar scene of the previous one (i.e. the camera has moved). The pose of the second camera with respect to the first one is define by the transformation:

$$T = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} \tag{3.42}$$

where R is the rotation matrix and t is the translation vector. This paragraph want to demonstrate the relation between this physical parameters expressing the motion of the camera, with the 2D-to-2D projective transformation which relate feature points in two correspondent images.

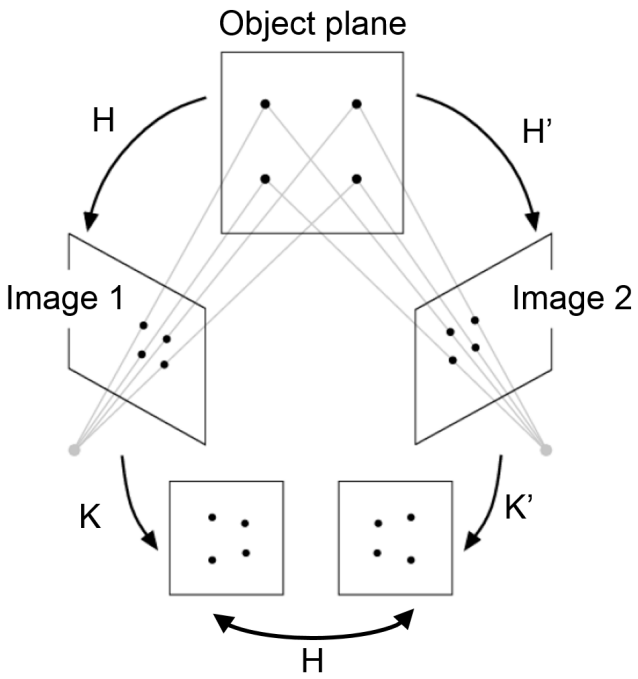


Figure 3.13: Scene planes and homographies relations.

As shown in the Table 3.1 the planar homography contains inside both the internal and external camera parameters and could provide the information about the position of the camera in the world reference frame. The 2D-to-2D homography on the other hand is linearly obtained from homogeneous 2D coordinate in $\mathbb{P}^2$ expressed in pixel and don't express inside any information related to the camera

sensor and pose of the image. For this reason, $H_{x\prime\to x}^{DLT}$ is called 2D uncalibrated homography G.

Knowing the camera calibration matrix K is possible to obtain an homography that relates points in calibrated coordinate: $H = K^{-1}GK$ At this point this calibrated homography can be decomposed to find the rigid (Euclidean) transformation R and t between the two cameras. It's important to underlie that this decomposition is not so straightforward as the 2D-to-2D uncalibrated homography could not be compatible with another concept of multiple view geometry: the epipole constrain.

Table 3.1: Homographies

3D-to-2D projective transformation	**Plane induced homography**	**2D-to-2D projective trasformation**
General 3D scene in CCS $\widetilde{x} = P_{3\times4}X_{cam}$ $P = K[I \mid 0]_{3\times4}$ $\widetilde{x} = K\,[I \mid 0]\,X_{cam}$	Planar 3D scene in CCS $\widetilde{x} = H_{3\times3}X_{cam}$ $H = K[I \mid 0]_{3\times3}$ $\widetilde{x} = K[i_1\ i_2\ 0]_{cam}$	Two images acquiring the same scene $x' = Hx$ $H_{x\prime\to x}^{DLT}$ $x = (u, v, w)$ $x' = (u', v', w\prime)$
General 3D scene in WCS $\widetilde{x} = P_{3\times4}X_{world}$ $P = K[R \mid t]_{3\times4}$ $\widetilde{x} = K\,[R \mid t]\,X_{world}$	Planar 3D scene in WCS $\widetilde{x} = H_{3\times3}X_{world}$ $H = K[R \mid t]_{3\times3}$ $\widetilde{x} = K[r_1\ r_2\ t]X_{world}$	

A planar surface in a 3D scene induces a unique projective transformation, called **planar induced homography**, that relates the projections in two views of any point belonging to a plane. The epipolar geometry between the two images could be seen as a point transfer via plane geometry which involves that since the set of 2D points in both images are projective equivalent to the 3D planar point set X, so they are also projective equivalent between each other. Mathematically:

$$\mathbf{x} = H\mathbf{X} \qquad \text{and} \qquad \mathbf{x}' = H'\mathbf{X}$$

The point $\mathbf{x}$ in the first image is forward projected to meet the plane π in $\mathbf{X} = H^{-1}\mathbf{x}$. This point is then backprojected in the image plane of the second image $\mathbf{x}^{'} = H'\mathbf{X}$. The composition of the two homography is an homography between the two images:

$$\mathbf{x}' = H'H^{-1}\,\mathbf{x} = H\mathbf{x}$$

$$\mathbf{x}' = H\mathbf{x}$$

So, if a plane is given (defined by his normal vector) is possible to express the homography induced by the plane as follow: Given the camera matrix for two

views acquiring the same planar scene: $P = K[I \mid \mathbf{0}]$ and $P^{'} = K[R|\mathbf{t}]$, and given a plane $\pi = (\mathbf{n}^T, d)^T$ it is possible to parametrize a point $\mathbf{X}$ as a ray that intersect the plane π with the condition $\pi^T \mathbf{X}_\pi = 0$, with point $\mathbf{X}_\pi = (\mathbf{x}^T, \rho)^T$.

$$\pi^T \mathbf{X}_\pi = 0 \qquad (\mathbf{n}^T, d)^T \begin{pmatrix} \mathbf{x} \\ \rho \end{pmatrix} = 0 \qquad \mathbf{n}^T \mathbf{x} + d\rho = 0 \qquad \rho = -\frac{\mathbf{n}^T}{d\mathbf{x}}$$

Then $\mathbf{X}_\pi = (\mathbf{x}^T, -\frac{-\mathbf{n}^T}{d}\mathbf{x})^T$

$$\mathbf{x}' = P'\mathbf{X}_\pi = K[R \mid \mathbf{t}]\mathbf{X}_\pi = K[R \mid \mathbf{t}](\mathbf{x}^T, -\frac{\mathbf{n}^T}{d}\mathbf{x})$$

$$\mathbf{x}' = KRK^{-1}\mathbf{x} - K\mathbf{t}K^{-1}\frac{\mathbf{n}^T}{d}\mathbf{x}$$

$$\Rightarrow H = K\left(R - \mathbf{t}\frac{\mathbf{n}^T}{d}\right)K^{-1}\mathbf{x} \tag{3.43}$$

Determining this homography allows to compute the following parameters:

- the normal $\mathbf{n}$ to plane;
- the rotation $\mathbf{R}$;
- the ration $\frac{t}{d}$ of the translation to the distance of the plane.

This is not an unique homography but a three-parameter family of homographies. From what has been shown before, since the point correspondences arise from images of the scene plane, they obey also to the Epipolar geometry i.e. $x_i'^T F x_i = 0$. This means that the homography H is compatible with F. To summarize, there are two relation between two views acquiring the same planar scene, first the Epipolar geometry and second the homography relation:

$$x \longleftrightarrow x' = Hx$$

And

$$x'^T F x = 0$$

Then:

$$(Hx)^T F x = 0 \qquad x^T H^T F x = 0 \tag{3.44}$$

This is true for all points x. Unfortunately, the estimation of F from point correspondences is degenerative in this case and we are not able to compute it. To assure that an homography H is compatible with a fundamental matrix F one must verify that the matrix $H^T F$ is skew-symmetric with $H^T F + F^T H = 0$.

3.2.1.7 Homography Decomposition

It has been demonstrated how a plane induced homography between two view can be parametrised in order to express the camera transformation and the inducing plane (H is function of R, $\mathbf{t}$ and $\mathbf{n}/d$. It has been also demonstrate as from a compatible H to F is possible to decompose the homography in a straightforward way and obtain a compatible configuration. Unfortunately, retrieve the real configuration is more complicate as there is an ambiguity in the solution. The problem to recovering the possible configurations of the cameras from a compatible homography is known as **the homography decomposition problem.** In the particular case of a calibrated camera moving frame by frame it turns out that there are only finite number of possible decomposition and adding some external information by some geometric assumption is possible to extract the correct camera configuration. Faugueras and Lustman in [37] describes the decomposition methods used in the present work. Other methods are used in [38] and [39].

3.2.1.8 Visual Odometer

As already said in the introduction to this chapter, one of the most important task for which computer vision algorithm has been developed is the Visual Odometry (VO). This is the process of estimating the ego-motion of a vehicle, human, robot and more, using the information contained in the frames acquired by a camera sensor. In this work the monocular visual odometry problem (information from a singular camera with scale factor unknown) is addressed. Monocular visual odometry operates by incrementally estimating the camera pose through examination of the changes that motion induces on the images acquired on each step. The introduced estimation error in each frame position accumulate over time and generate a drift from the real path in the estimated trajectory [40]. This drift can be reduced optimizing the frame-to-frame pose overt the last cameras or combining other sensor measurements in the solution estimation (GNSS, LiDAR , IMU) [41] [42]).

Assuming a set of images I_k acquired sequentially at time k, between two subsequent frame there is a three-dimensional rigid transformation T_k that represent the motion of the body.

$$T_k = \begin{bmatrix} R_{k,k-1} & \mathbf{t}_{k,k-1} \\ 0 & 1 \end{bmatrix} \tag{3.45}$$

This transformation brings the camera center coordinate of the first image to the camera center coordinate of the second image:

$$C_n = C_{n-1} T_n \tag{3.46}$$

Or:

$$\mathbf{x}'_k = \lambda(R_{k,k-1}\mathbf{x}_k + \mathbf{t}_{k,k-1}) \tag{3.47}$$

Taking into account equation 3.8, the relation expressed in 3.47 becomes

$$\mathbf{x}'_k = K'R_{k,k-1}K^{-1}\mathbf{x}_k + \mathbf{t}_{k,k-1}/d) \tag{3.48}$$

where d is the distance between the coordinates of object point in the real word and the coordinates of the camera center. Knowing d means solving the scale problem. It can be done mainly in three ways:

- stereo camera approach: knowing the pose of the two cameras and the baseline between them give the distance d by triangulation;
- monocular camera approach: knowing a-priori information of the object (size, coordinates) allows to compute the distance d;
- monocular camera approach: using external sensors to measure the distance d;
- monocular camera approach: camera facing down with fixed height gives the Z coordinate of the camera center and consequently the d value by geometry relations.

The requirements to apply the equations previous described, for Two-View Geometry estimation, are the capability to identify, extract and match the best possible features from the frames. This concept brings to the following assumption for the efficiency of the visual odometry procedure:

- sufficient illumination in the environment;
- dominance of the static scene over moving objects;
- enough texture to allow apparent motion to be extracted;
- sufficient scene overlaps between consecutive frames.

Under this assumption is possible to perform feature extraction and matching with state of art procedures [33], [43], [44]. Then the estimation of the relative camera pose can be performed with the decomposition algorithm previously presented. As reported in [40] the complete VO algorithm is the following:

- capture new frame I_k;
- extract and match features between I_{k-1} and I_k;
- compute essential matrix for image pair I_{k-1}, I_k;

- decompose essential matrix into R_k and t_k, and form T_k;
- compute relative scale and rescale t_k accordingly;
- concatenate transformation by computing $C_k = C_{k-1}T_k$;
- repeat the procedure.

3.2.2 The photogrammetric approach

In the last years, photogrammetry has becomes enormously important in numerous applications, different from the survey for which is born, thanks to the digitalization and miniaturization of the camera sensors and the automatization of the computation process. "Photogrammetry is the science of obtaining reliable information of physical objects and environment through registration, measurement and interpretation processes of photographic and digital images formed by electromagnetic radiant energy and other physical phenomena." [45]. The definition intend to highlight the photogrammetry as a set of techniques based on the analysis and processing of digital or analog photographs (or frames) that, to a first approximation, can be considered as a central perspective (as shown in Figure 3.14).

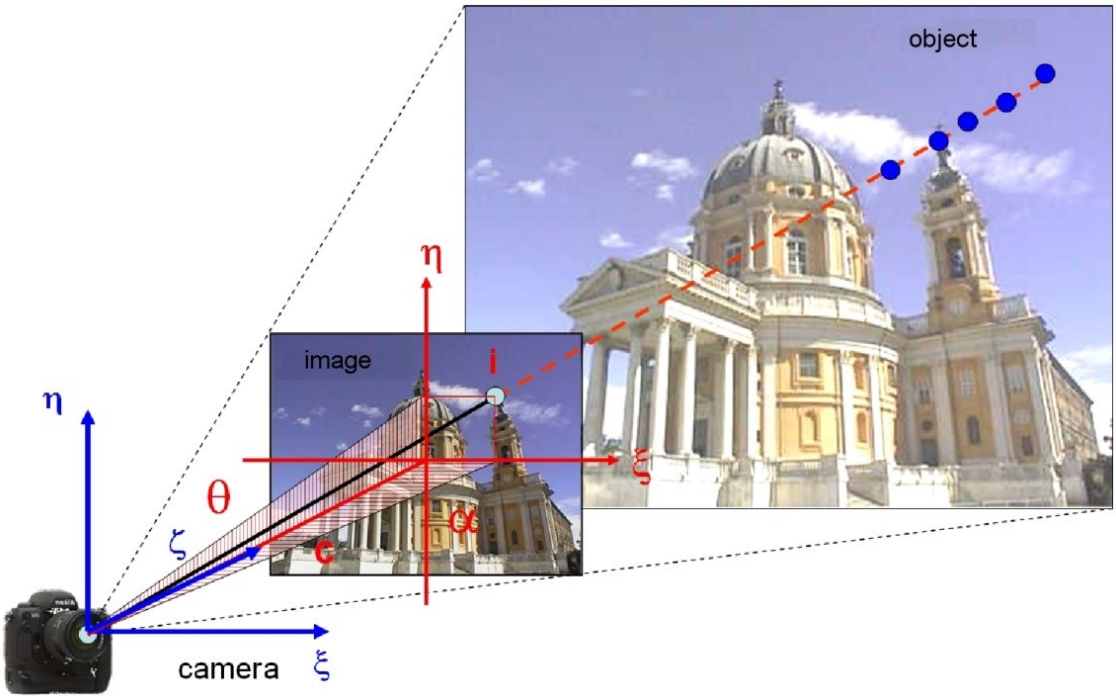


Figure 3.14: Example of central perspective.

3.2.2.1 Collinearity equation

To reconstruct the position and the shape of the objects from photographs one must know the geometric relationships according to which the images are formed. The cameras used in photogrammetry produce frames that, in first approximation, can be considered central perspective, geometrically strict, of the object. When a frame is acquired, the object point P, the shooting center O and the image point P' lie on a same straight line, as shown in Figure 3.14. Suppose to introduce a new land system X', Y', Z', parallel and rotated in the space to the image system ξ, η, ς ($\varsigma = 0$ for the image points and $\varsigma = c$ for the shooting center), but with the origin coincident to the system X, Y, Z. The collinearity conditions that express the alignment between the points in the system X', Y', Z' can be expressed in the following way:

$$\frac{\xi - \xi_0}{c} = \frac{X' - X'_0}{Z'_0 - Z'} \tag{3.49}$$

$$\frac{\eta - \eta_0}{c} = \frac{Y' - Y'_0}{Z'_0 - Z'} \tag{3.50}$$

The coordinates X', Y', Z' of the point P and the coordinates X'_0, Y'_0, Z'_0 of the projection center may be related in the system X, Y, Z by the spatial rotation matrix R:

$$\begin{pmatrix} X - X_0 \\ Y - Y_0 \\ Z - Z_0 \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \cdot \begin{pmatrix} X' - X'_0 \\ Y' - Y'_0 \\ Z' - Z'_0 \end{pmatrix} \tag{3.51}$$

So, multiplying the matrix above for the matrix $R^T = R^{-1}$, and replace in the equations, explicit the image coordinates, the relationship between the image coordinates and the ground ones are obtained:

$$\xi = \xi_0 - c\,\frac{r_{11}(X - X_0) + r_{21}(Y - Y_0) + r_{31}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} = \xi_0 - c\,\frac{Z_x}{N} \tag{3.52}$$

$$\eta = \eta_0 - c\,\frac{r_{12}(X - X_0) + r_{22}(Y - Y_0) + r_{32}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} = \eta_0 - c\,\frac{Z_y}{N} \tag{3.53}$$

Because the system X', Y', Z' is parallel to the image system ξ, η, ς, the terms r_{ik} represent:

- the cosines of the angles between the coordinate axes of the image system and the ground system, or
- the functions of the angles ω, φ, k of which the frame was rotated, respect to the ground system, when the frames were taken.

Now is possible to express the collinearity equations from image coordinates to object coordinates, as follows:

$$X = X_0 + (Z - Z_0) \frac{r_{11}(\xi - \xi_0) + r_{12}(\eta - \eta_0) - r_{13}c}{r_{31}(\xi - \xi_0) + r_{32}(\eta - \eta_0) - r_{33}c} \tag{3.54}$$

$$Y = Y_0 + (Z - Z_0) \frac{r_{21}(\xi - \xi_0) + r_{22}(\eta - \eta_0) - r_{23}c}{r_{31}(\xi - \xi_0) + r_{32}(\eta - \eta_0) - r_{33}c} \tag{3.55}$$

In conclusion, any object point corresponds to an image point while to each image point may correspond infinite object points, due to the presence of the Z to the second member. This explains why it is impossible to reconstruct the spatial geometry of an object from only one frame. In the paragraph Two-view geometry, this problem has been be presented.

3.2.2.2 The Syntetic Solid Image

A Solid Image or and RGB-D image (Red,Green,Blue and Distance channel) (Figure 3.15) is a classical RGB digital image with known internal and external orientation parameters, where a distance between the projection center and the acquired objects are recorded for each pixel. Therefore, distance values are stored in an additional matrix with the same pixel size, number of columns, and number of rows as the RGB matrix.

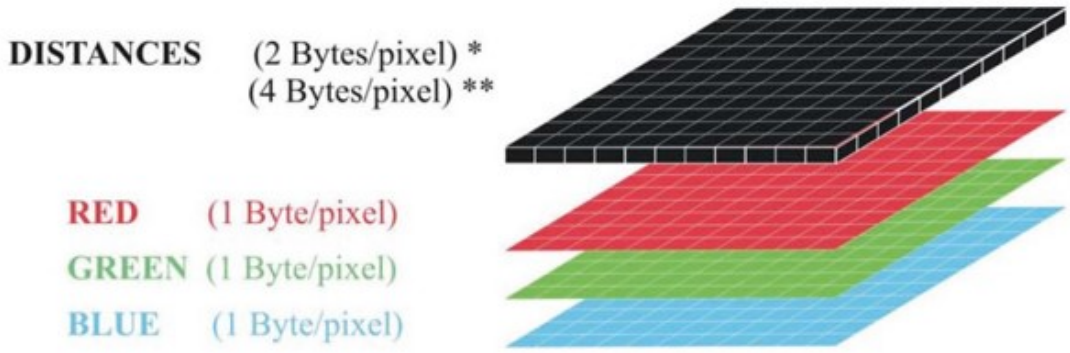


Figure 3.15: Solid Image structure.

To generate a RGB-D image the following information are needed:

- the external orientation parameters corresponding to the position and orientation of the camera $(X_0, Y_0, Z_0, \omega, \phi, \kappa)$;
- the internal orientation parameters corresponding to focal length, the principal point position of the camera (c, ξ_0, κ_0) and distortions (the generated images are synthetic and are considered without distortion);

- the number of pixels in the columns and the rows of Solid Images $(nrow, ncol)$ and the image pixel size d_{pix}.

The external orientation parameters and the distance information can be obtained from any a-priori information of the object acquired. It can be an active sensor which measure directly the distance from the image plane to the object or a realistic 3D model of the area of interest, with both geometric and color information, obtained from a LiDAR survey. The internal orientation parameters can be selected by the user to generate a synthetic camera or taken from a calibrated one. The interest in this kind of image is that gives all the information useful for applying collinearity equations in a multiple-view geometry problem. Moreover, this kind of image can be automatic generated from a 3D model in a fast way to populate a database with all the information of a real object scene. These information are not only pixel colors and distance but also the exact position in the real word of this synthetic image.

3.2.2.3 The IRBL positioning

Let's imagine an user acquiring an image of a real 3D world object already documented by a RGB-D database of images. If through any image retrieval procedure one is able to extract the correct correspondent image from the database previous described, is possible to estimate the position and the attitude of the user camera applying simply the collinearity equations. In fact from any central projection, if the internal orientation parameters (ξ_0, η_0, c) are known is possible to define a direction in the space as:

$$\alpha = arctan \frac{\eta_I}{\sqrt{c^2 + \xi_i^2}} \tag{3.56}$$

$$\theta = arctan \frac{\xi_I}{c} \tag{3.57}$$

Knowing this direction means knowing the line on which the 3D object point belong. There are infinite possible points reprojected in the image plane. The correct 3D point can be retrieved from the distance information stored in the RGB-D image pixel:

$$X_0 = d \cos\alpha \sin\theta \tag{3.58}$$

$$Y_0 = d \sin\alpha \tag{3.59}$$

$$Z_0 = -d \cos\alpha \cos\theta \tag{3.60}$$

Once all the 3D object points are computed from a RGB-D image with previous procedure the information of the real word can be used for any new image acquired. So, if an user acquire a new image from a different position, is possible to apply the collinearity equations (Eq. 3.55) and estimate position and attitude of the camera.

3.3 GNSS positioning with smartphones

In recent years the surprising increase in computing, memory and data connection capabilities of mobile phones has allowed the spread throughout the world of these smart devices, making them indispensable tools not only for communication but also for carrying out very complex tasks. Therefore are not surprising the data appeared in the Global Digital 2018 report [46], which states that in the world, the number of connected smartphones is about 5.13 billion. All these users have access to a powerful technology rich of micro-electromechanical systems (MEMS) sensors inside it, able to acquire huge amount of data and information continuously, in every moment of the day, in every corner of the world. These sensors are RFID, cameras sensors, GNSS chipset, Inertial Measurement Unit (IMU) platforms, barometer, altimeter and more, that empower customers to plan their activities, to interact with others, to share their moments, to navigate in unknown environments.

In this chapter, smartphone technology and its relation with the field of navigation and positioning will be analysed. It has been already shown in section 3.2.1 how the camera sensor installed in new generation smartphones can be a powerful tool for acquiring image data to process in computer vision algorithms and retrieve positioning information. The focus in this section will be on the GNSS positioning with smartphone. The GNSS chipset installed on new generation devices is able to provide positioning coordinates in global reference system (the other sensors, like IMU, in local frames) while other sensors like inertial system can also monitor the human body motion. State of art research have demonstrated that both pedestrian and vehicle navigation can be performed very well with the Assited GPS (A-GPS) algorithms and the cellular based station triangulation.

As well known, the GNSS positioning is only useful in outdoor scenario because it needs the satellite visibility. When outdoor scenarios are considered, smartphone technology can provide positions with a quite good level of accuracy also thanks to the aiding of correction models and clock data. These methods have improved ranging accuracy to some meters, leaving the dominant sources of error in current consumer devices dependent from multipath and front-end-noise-induced. Under good multipath conditions, 2-to-3-meter-accurate positioning is typical; under adverse multipath, accuracy degrades to 10 meters or worse. This is the typical case of an urban environment, where an user can be inside a buildings or in urban canyons.

Numerous research have been made in order to obtain better performances in these unfavourable environments and to reach sub-metric level of accuracy with smartphone devices. Unfortunately, to apply signal processing algorithms and post-processing procedure the raw data acquired by the front end must be used. This was almost impossible till some years ago as the smartphone developers allowed to extract directly the PVT solution from their chipsets. Fortunately, in May 2016, Google announced the realise of a new Operating System (OS), the Android Nougat 7.0, from which the GNSS raw data extraction is possible. This announcement

caused a resumption of research in the field of mobile positioning and opened the way to the attempt centimeter positioning. Moreover, on September 2017, Broadcom announced the world first mass-market, dual frequency GNSS receiver device, the BCM47755. This strong technology innovation can also open new scenarios in the field of positioning and navigation with smartphones.

The following section will presents basic concepts of GNSS positioning with smartphone. The procedure used for perform positioning are the ones used in geodetic application so the focus will be on the structure of these smartphone information and how to access to these raw data with the new Android OS.

3.3.1 Principles of GNSS positioning

The GNSS positioning relies on distance measurements (pseudoranges) between a receiver antenna and a constellation of satellites in orbit around the earth. Thus, knowing one satellite position, the receiver position lies somewhere on a sphere around the satellite with a radius equal to the range. If the range measurements of three satellites are available, the three spheres intersect at two points. Since one point is not located close to Earth surface, the second point is the true position of the receiver. Usually at least four satellites of the same constellationare needed to account for the time difference between satellites clocks and receiver clock [47] [48].

There are currently four GNSS constellations in operation or in deployment phase: GPS (USA), GLONASS (Russia), BeiDou (China) and Galileo (Europe), that are integrated by different augmentation systems, at local or global scale. For estimating the geometric range, all GNSS signals are modulated with a Pseudo Random Noise (PRN) code, a few milliseconds long sequence that allows to uniquely identify the satellite that transmits the signal. The receiver is able to reply the same signal in order to compare and align the replied copy of the PRN code with that received by the satellite. The distance between receiver and satellite is proportional to the signal propagation time, if the transmitter and receiver clocks are perfectly synchronised.

GNSS signals are affected by errors, including thermal noise, uncompensated biases, multipath, and propagation effects. In particular, atmospheric propagation is the biggest error source, inducing ionospheric and tropospheric delays and ionospheric scintillations. If these biases are not estimated and compensated, the positioning error can achieve up to 50 m, making the GNSS positioning useless for most of applications. In order to obtain a position fix, three different techniques can be followed: to differentiate, to model or combine the so called GNSS observables, pseudoranges and carrier phase measurements (Eq. 3.62). Considering this last kind of measurements, another variable affects the quality of positioning: the phase ambiguity, indicated with N in Eq. 3.62, that must be estimated as integer value if a centimetre level of accuracy is required. The range observation are

expressed as:

$$P_k^j(i) = \rho_k^j - cdT_k + cdt^j + \alpha_i I_k^j + T_k^j + m_{i,k}^j + E_k^j + cL_k^j + \varepsilon_{i,k}^j \quad (3.61)$$

$$\phi_k^j(i) = \rho_k^j - cdT_k + cdt^j - \alpha_i I_k^j + T_k^j + Mi_k^j + E_k^j + \lambda_i Ni_k^j + \varepsilon_k^j \quad (3.62)$$

where $P_k^j(i)$ and $\phi_k^j(i)$ represent the pseudorange and carrier-phase measurements on the i-th frequency, respectively. ρ_k^j is the geometric range between the satellite j and the receiver k, cdT_k and cdt^j ate the biases related to receiver and satellite clocks multiplied by the speed of light. The ionospheric propagation delay is represented by $\alpha_i I_k^j$ (with a known coefficient$\alpha_i = f_1^2/f_i^2$ that depends on the i-th frequency). T_k^j is the tropospheric propagation delay, Mi_k^j is the multipath error, E_k^j is the ephemeris error. The carrier-phase ambiguity multiplied by the frequency length is $\lambda_i Ni_k^j$ and, finally, the random errors ε_k^j. As widely described in literature, GNSS positioning can be realized adopting two different approaches: using post-processing techniques or real time methods. Post-processing techniques are usually focused when a high level of accuracy is required or when it is not possible to estimate and to apply a model of biases in real time. This generally happens considering the typical receivers used for positioning purposes, such as geodetic or GIS ones [49].

With the advent of mass-market devices, GNSS positioning has become more common because the cost of GNSS receivers and antennas has been decreased up to few US dollars. This has made possible to insert these chipset into mobile phones also for positioning and geo-localization, even if some limitations persist. Most GNSS chipsets installed inside smartphones or tablets are single-frequency receivers: it means that they provide only measurements related to one frequency (the L1 band). In these cases, it is not possible to apply the double or triple differences approach [50] or to combine different observations [51]. So, the two possible approaches are to compute single differences (considering one receiver and a reference satellite) or to model the GNSS biases (e.g. Iono and tropospheric delays, satellite and receiver clock drifts) using mathematical models.

3.3.2 Assisted GNSS

The spread of GNSS positioning in smartphones has been pushed mainly from the use of Assisted GNSS (A-GNSS) techniques [52]. A-GNSS is based providing certain information that the GNSS receiver would ordinarily have received from the satellites themselves, through an alternative communication channel. A-GNSS is then a complementary technology, which enables two key aspects of GNSS for smartphones: speed and sensitivity. A-GNSS reduces the time required to perform a position fix, i.e. the Time To First Fix (TTFF) [52]. GNSS receivers designed for geodetic, survey and military purposes have no strict requirements in terms of TTFF. This is the reason why GPS was originally developed to provide a fix

in around 1 minute of processing. It is obvious that when moving to the mobile world, the TTFF must significantly decrease down to few seconds as maximum to guarantee the usability. At the same time, open-sky and ideal propagation conditions, which are typically available for certain GNSS professional applications, are usually not fulfilled when moving to the mass-market. Receivers for phones must deal with signals received at lower power and quality. The two main objectives of A-GNSS are to reduce the TTFF and to improve the receivers sensitivity.

The core idea is to provide an aid to the device via a wireless network. Such aid includes the content of the navigation message (ephemeris, almanacs, ionospheric corrections), which can then been obtained in a few milli-seconds, rather than demodulating a full navigation message. In addition, advanced information can be downloaded, such as precise ephemeris, approximate user position and time, precise ionospheric corrections, and estimated signals acquisition parameters (estimated Doppler shift, estimated delay, navigation message). The Systems Of Chip (SOC) inside the smartphones integrate the entire GNSS technology in a single device and manage all the previous data to provide Position, Velocity and Time (PVT) solution in National Marine Electronics Association (NMEA) format or as direct information for the user application. Then, the OS manage the calls for data processing and sensors integration. Considering the Android environment, when an application requests location update without accessing raw measurements, it performs some callback from the different chipset of the microprocessor in charge of acquiring GNSS, cell or Wi-Fi information (NETWORK_PROVIDER call). As the SOC is a black box, the user can directly control just the access to the fine or coarse location. This means that the NETWORK_PROVIDER call can be enabled or disabled and the positioning can move from the assisted to the GNSS only. The developers instead, are able to manage the workflow to retrieve the user location i.e. when and for how long the chipsets should listen and update the information. Moreover, the new smartphone generation integrates numerous motion and orientation sensors for cross aiding with A-GNSS. The Android sensor framework provides several classes and interfaces able to determine the performances, the minimum rate, the maximum range, the power requirements and more.

3.3.3 The GNSS raw measurements architecture

Accesing to the raw data acuqired by the GNSS chipset could gives several advantages increasing positioning performances and allowing the use of more advanced processing techniques that till now have been restricted only to geodetic receivers. During Google I/O 2016, Android experts announced the possibility to access raw data from GNSS chipsets inside smartphones. On the Andorid developer website [53], the list of smartphones providing GNSS raw measurements is continuously updated. Information also include the latest available Android version, the presence of an Automatic Gain Control (AGC), the availability of both the

decoded navigation message and of the phase measurements (accumulated delta range), information about the clock, and the support of multi-constellation and multi-frequency signals. The first available devices running the Android N OS were the Nexus 9 tablet, Nexus 5x phone, Nexus 6p phone, Pixel phone and the Pixel XL phone. The Nexus 9 was the only one able to provide dual constellation (GPS and GLONASS) pseudorange and carrier-phase measurements. Although the Nexus 5x, Nexus 6p and Pixel phones supported GPS and GLONASS, only GPS raw pseudorange measurements were available. Furthermore, the Nexus 9 tablet was the first device with duty cycling disabled, starting from Android N 7.1 release. Therefore, it is suitable for collecting continuous carrier-phase measurements over periods of many minutes.

The Android OS is composed by packages of APIs, which are a collection of protocols allowing users to access the system's functionalities. The primary API that manage the location framework in android devices is the Fused Location Provider (FLP), available in **com.google.android.gsm.location** package, which combines GNSS, Wi-Fi, cell, accelerometer, gyroscope and magnetometer in order to provide an integrated PNT solution. This solution is used by all the google app environment to show the position (Google Maps) or to pop-up location related information (Google Event). The FLP manage autonomously which sensor to turn on based on the accuracy and frequency requirements of the app developer. The inner sensor-fusion algorithm developed by Google have to satisfy the huge panorama of final users which requirements are usually different from the ones of the best accuracy in positioning. In particular, Google software engineers put their effort on find the best balance between coverage, accuracy, latency and power consumption. The algorithms implemented can be considered as a black-box for the users.

Another package which contains location-based API is **com.google.android.-location** which manage all the classes related to the GNSS chipset. This API was firstly released in API level 1, the base framework of the android operating system without allowing the access to the source code and to the measurements acquired by the chipset. Numerous classes and methods have been added during the various OS release till 2016 when Android 7.0 Nougat and the related API level 24 introduces the direct access to the classes that provides the raw data measurements of the GNSS chipset. Until API level 23, the access was limited to C/No, azimuth, NMEA sentences, some satellite elevation and PVT solution with related timestamp. All this usually referred to the GPS constellation alone.

From API Level 24, developers have access to numerous classes that extract raw and computed GNSS information directly from the chipset. Although the recent features introduced, Google push the developer's community to use the FLP framework for developing location and context application. On the other hands, the great interest raised from the new location API in the GNSS community, has pushed Google and the OEMs to invest more and more in the development of hardware and software for positioning. The direct access to the raw data in fact allows to apply

directly positioning and navigation algorithms to the satellites measurements and to move the attention on the accuracy of the positioning and navigation solution more than on the performances of the smartphone. Figure 3.16 shows the software architecture comparing the two packages provided by Google:

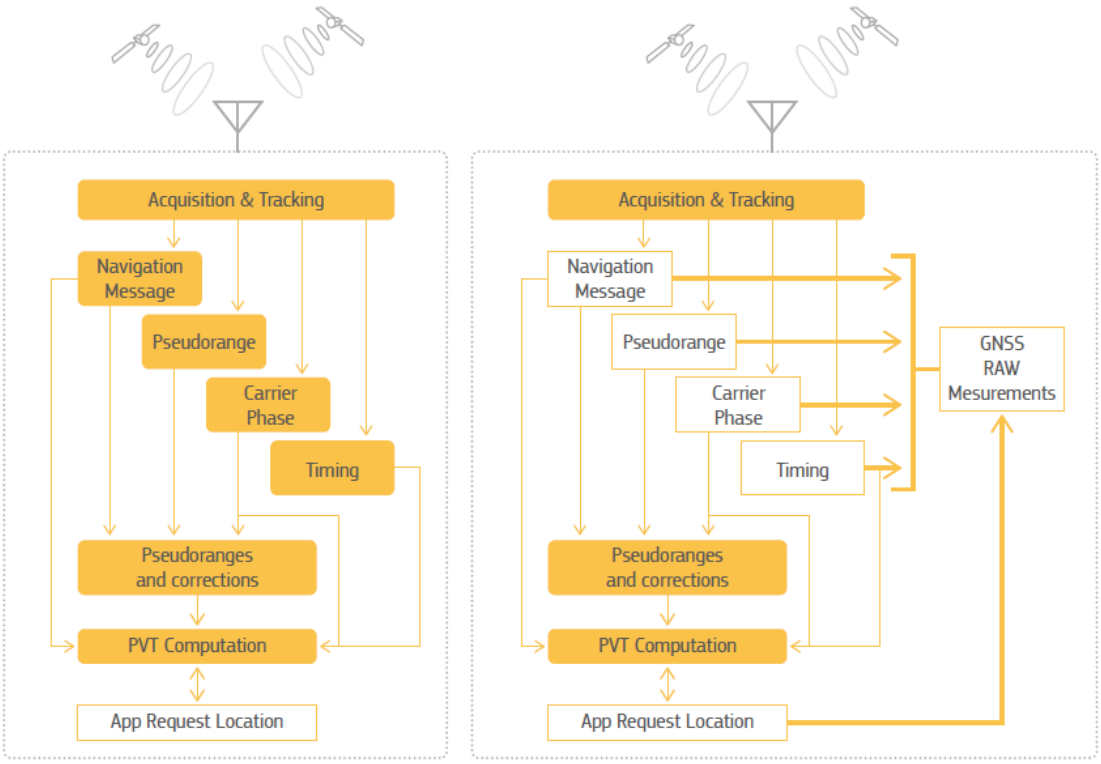


Figure 3.16: android.gsm.location package vs android.location package after API level 24. Ref: GSA GNSS Raw measurement task force – White paper:"Using GNSS raw measurements on android devices".

In order to use raw measurements for testing and post-processing procedure, GNSSs experts needs to convert data fields provided by the API classes into standard format readable from state-of-art processing software. Unfortunately, the understanding of the measurements obtained by the different location classes is not straightforward, java developer are usually unfamiliar with the details of the GNSSs and vice versa positioning expert are not totally aware of the API structure. In this chapter will be described the most important data fields obtainable from the location API and how from this is possible to obtain GNSS data in RINEX format and navigation message in NMEA format.

In the Android Developers webpage, there is a list of the classes in the android.location API with a brief description of the services that each one provides

[54]. The two most important classes from which all the interesting raw data are extracted are:

- **GnssClock**: the class contains a GPS clock timestamp;
- **GnssMeasurement**: the class represents a GNSS satellite measurement, containing raw and computed information.

The data fields obtained from these two classes are the primary information from which is possible to compute a PNT solution. Table 3.3.3 describes the data fields of the raw measurements.

Table 3.2: Description of GNSS raw measurements

Raw data field	**Description**
ElapsedRealtimeMillis	Returns milliseconds since boot, including time spent in sleep
TimeNanos	GNSS receiver internal hardware clock value in nanoseconds. This value is expected to be monotonically increasing while the hardware clock remains powered on
LeapSecond	The leap second associated with the clock's time
TimeUncertaintyNanos	The clock's time Uncertainty (1-Sigma) in nanoseconds
FullBiasNanos	The difference between hardware clock getTimeNanos() inside GPS receiver and the true GPS time since 0000Z, January 6, 1980, in nanoseconds. This value is available if the receiver has estimated GPS time. If the computed time is for a non-GPS constellation, the time offset of that constellation to GPS has to be applied to fill this value
BiasNanos	The clock's sub-nanosecond bias
BiasUncertaintyNanos	The clock's Bias Uncertainty (1-Sigma) in nanoseconds
DriftNanosPerSecond	The clock's Drift in nanoseconds per second. A positive value indicates that the frequency is higher than the nominal (e.g. GPS master clock) frequency

Table 3.2: Description of GNSS raw measurements

Raw data field	**Description**
DriftUncertaintyNanosPerSecond	Uncertainty of DriftNanosPerSecond
HardwareClockDiscontinuityCount	Count of hardware clock discontinuities
Svid	Satellite ID
TimeOffsetNanos	The time offset at which the measurement was taken in nanoseconds. The reference receiver's time from which this is offset is specified by getTimeNanos()
State	Gets per-satellite sync state
ReceivedSvTimeNanos	The received GNSS satellite time, at the measurement time, in nanoseconds
ReceivedSvTimeUncertaintyNanos	Gets the error estimate (1-sigma) for the received GNSS time, in nanoseconds
Cn0DbHz	Carrier-to-noise density
PseudorangeRateMetersPerSecond	Gets the Pseudorange rate at the timestamp in m/s
PseudorangeRateUncertainty MetersPerSecond	Uncertainty of PseudorangeRateMetersPerSecond
AccumulatedDeltaRangeState	It indicates whether getAccumulatedDeltaRangeMeters() is reset or there is a cycle slip (indicating 'loss of lock')
AccumulatedDeltaRangeMeters	The accumulated delta range since the last channel reset, in meters
AccumulatedDeltaRangeUncertainty Meters	Uncertainty of AccumulatedDeltaRangeMeters
CarrierFrequencyHz	Carrier frequency at which codes and messages are modulated
MultipathIndicator	Gets a value indicating the 'multipath' state of the event
SnrInDb	Gets the (post-correlation and integration) Signal-to-Noise ratio (SNR) in dB
ConstellationType	Constellation type
ConstellationType	Gets the Automatic Gain Control level in dB

The official Android application developed with the aim of provide all the raw data described above is called GnssLogger [53]. It logs the measurement data in

comma-separated-value (csv) text format (Figure 3.17). Other third part applications utilizing GNSS raw measurements has been developed in the last years. A list of devices capable of providing raw measurements data is maintained at [53]. The class android.location.GnssNavigationMessage allows to receive GNSS satellite Navigation Message from the GNSS engine. The Gnsslogger app logs the decoded ephemeris data in decimal representations of the bytes defined by the respective constellation Interface Control Documents (ICDs). The data field with relative description are reported in the documentation of Android Developer.

```
1  #
2  # Header Description:
3  #
4  # Version: v2.0.0.1 Platform: 8.1.0 Manufacturer: Xiaomi Model: MI 8
5  #
6  #
   Raw,ElapsedRealtimeMillis,TimeNanos,LeapSecond,TimeUncertaintyNanos,FullBiasNanos,BiasNanos,BiasUncertain
   tyNanos,DriftNanosPerSecond,DriftUncertaintyNanosPerSecond,HardwareClockDiscontinuityCount,Svid,TimeOffse
   tNanos,State,ReceivedSvTimeNanos,ReceivedSvTimeUncertaintyNanos,Cn0DbHz,PseudorangeRateMetersPerSecond,Ps
   eudorangeRateUncertaintyMetersPerSecond,AccumulatedDeltaRangeState,AccumulatedDeltaRangeMeters,Accumulate
   dDeltaRangeUncertaintyMeters,CarrierFrequencyHz,CarrierCycles,CarrierPhase,CarrierPhaseUncertainty,Multip
   athIndicator,SnrInDb,ConstellationType,AgcDb,CarrierFrequencyHz
7  #
8  # Fix,Provider,Latitude,Longitude,Altitude,Speed,Accuracy,(UTC)TimeInMs
9  #
10 # Nav,Svid,Type,Status,MessageId,Sub-messageId,Data(Bytes)
11 #
12 Raw,941216,639146000000,,,-1222766447854106535,0.0,4.830576172461286,,,0,25,0.0,17466,466286910794362,24,
   26.72406005859375,-184.4775228376966,0.010578871704638004,1,-177593.7370110726,0.005289435852319002,1.176
   45005E9,,,,1,,6,,1.17645005E9
13 Raw,941216,639146000000,,,-1222766447854106535,0.0,4.830576172461286,,,0,26,0.0,17466,466286913722853,8,3
   0.06854248046875,376.1208804753842,0.007198055274784565,1,411452.32560526405,0.0035990276373922825,1.1764
   5005E9,,,,0,,6,,1.17645005E9
14 Nav,96,769,1,-1,5,43,-16,-128,0,0,24,24,0,0,86,104
15 Nav,97,769,1,3,5,43,-16,-128,0,0,24,24,0,0,86,104
```

Figure 3.17: Example of log txt file provided by the GnssLogger application, including GNSS raw measurements.

Deriving standard format

In order to produce Radio Technical Commission for Maritime Services (RTCM) or Receiver-Independent Exchange (RINEX) formats, the raw measurements provided by the smartphone receiver need to be managed, disposed and combined to obtain the data at the design of the standard require. For each satellite the identification code, the code phase, the carrier phase, the Doppler shift and the signal-to-noise ratio C/N_0 have to be reported together, along with the correspondent satellite identification and GNSS epoch. In the following paragraph the analytic computation to obtain this data from the smartphone raw measurements will be described.

1 Constellation and Satellite ID: The ConstellationType data fields provide an identification value for each constellation as reported in Table 3.3, while the Satellite ID is provided in the Svid field and follows the identification standards for the satellites.

Table 3.3: Constellation type values

Constellation Type	
STATUS	VALUE
CONSTELLATION BEIDOU	5
CONSTELLATION GALILEO	6
CONSTELLATION GLONASS	3
CONSTELLATION GPS	1
CONSTELLATION QZSS	4
CONSTELLATION SBAS	2
CONSTELLATION UNKNOWN	9

2 UTC Time: The Universal Time Coordinated (UTC) time of each epoch of observation is computed from the measured time of the GNSS chipset. Every time that a full set of observation is acquired, the time of the first observation is used to define the change in the epoch. As this time is provided in GPS time, this has to be converted in UTC time.

3 Pseudoranges: The pseudoranges measurement are not provided directly from the GNSS chipset but have to be computed from the difference between the time of the measurement at the receiver and the time of transmission of the signal:

$$\rho = (t_{Rx} - t_{Tx}) \cdot c \tag{3.63}$$

Where t_{Tx} is the received GNSS satellite time in ns at the measurement time, i.e. the GNSS reference time when the signal was transmitted, t_{Rx} is the measurement time and c is the speed of light in vacuum. These two times terms need to be reconducted to the same GNSS reference time. Figure 3.18 shows how t_{Tx} and t_{Rx} are obtained from the Android API and which classes need to be calls. If class getReceivedSvTimeNanos() return directly the GNSS satellite time in ns for t_{Rx} location.clock classes return values from which the measurement time can be reconstructed and reconducted to the GNSS time:

$$t_{RxGNSS} = TimeNanos + TimeOffsetNanos \tag{3.64}$$
$$-(FullBiasNanos(1) + BiasNanos(1)) \tag{3.65}$$

Where $TimeNanos$ is the time of the measurement from the switch on of the internal GNSS hardware clock. $FullBiasNanos$ is the difference between $TimeNanos$ inside the GNSS receiver and the true GPS time since 0000Z, 6 January 1980. As the time transmitted is provided for each GNSS system in his own time reference,t_{RxGNSS} must be aligned to each frame in order to obtain the correct pseudorange or alternatively all the t_{Tx} must be align to

the t_{RxGNSS}. The complete approach is provided by Google in [16] and Figure 3.18.

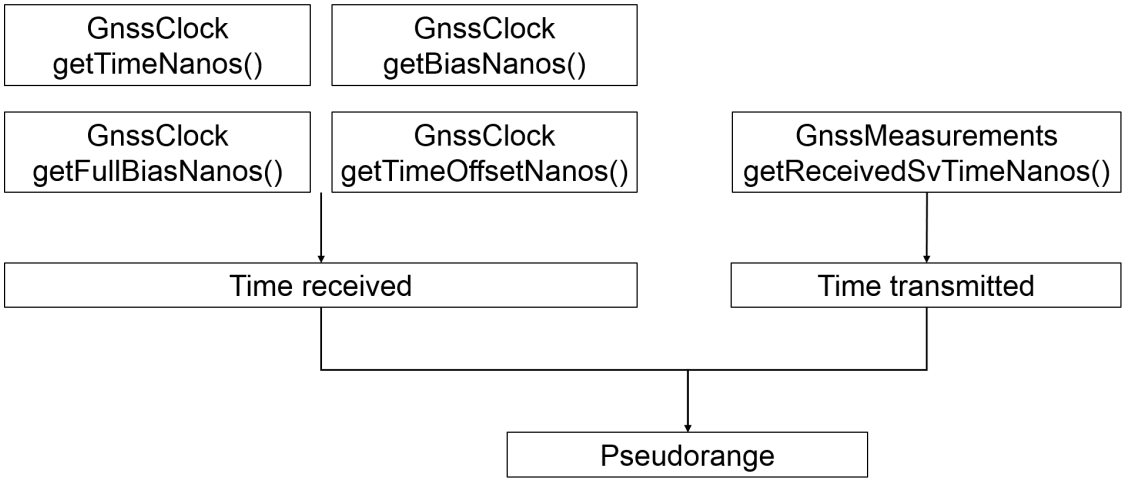


Figure 3.18: Flowchart of the different classes necessary to retrieve pseudoranges

4 Carrier Phase: The getAccumulatedDeltaRangeMeters() class returns the displacement of range from the last channel reset, in meters. At the end of a given interval of observation, a whole number of cycles will be counted usually together with a fraction of a cycle. This fractional phase is the carrier phase measurement:

$$CarrierPhase = -\frac{AccumulatedDeltaRange}{k} \tag{3.66}$$

If a cycle slip occur, the receiver loses this count, for this reason the class AccumulatedDeltaRangeState() provide a flag that assure the validity of the measurement. (ADR_STATE_VALID if VALUE=1)

5 Doppler shift: To obtain Doppler shift value, the line of sight velocity of the satellite v_{LOS} that the location API provide in the PseudoRangeRateMetersPerSecond data field is needed. Knowing the relations:

$$\lambda = \frac{c}{f_0}$$

$$f_D = \frac{v_{LOS}}{c} f_0$$

the Doppler shift is obtained as:

$$DopplerShift = -\frac{PseudoRangeRateMetersPerSecond}{k} \tag{3.67}$$

6 Signal to Noise ratio C/N_0: The C/N_0 is measured at the antenna input and provided inside the Carrier-to-noise density value (dBHz) which is obtained from the getCn0DbHz call.

3.4 Inertial Navigation Systems

Inertial Navigation Systems (INS) refers to a sensor platform able to sense the movement of a body and to process the acquired information to obtain position, velocity and attitude angles. It is composed by an Inertial Measurement Unit (IMU) and a central processing unit. The IMU are composed by accelerometers, which measure specific forces f and gyrometers, which measure the rate of rotation along specific axes ω. Inertial navigation is based on Dead Reckoning which means that the displacement of the body on which the INS is attached is computed by sensing how much it has move in a specific time period from the previous known position. This is accomplished by using the kinematic relation between acceleration, velocity and position. In particular, knowing acceleration measurements is possible to integrate once to obtain velocity and integrate once again to yield position. If one compute this displacement, then the direction of the movement is needed in order to extract the new position. An INS, where a stable platform is rigidly mounted on the body and move integrally with it is known as strapdown system. On the other hand, a system mechanically isolated from the host vehicle is known as gimballed. In function of their level of performance, INS suffers of instability in the processing which cause a fast drift to the position estimation. This well known drawback is usually compensate by hybridizing the INS architecture with a complementary technology, the GNSS [55]. The strong point of GNSSs is the long-term accuracy which compensate the fast error drift of the INS, while the strong point of the INS is the immunity from perturbation and the high update rate, which compensate the susceptibility of GNSS to interferences. Figure 3.19 express the navigation problem in INS where $\tilde{f}^p_{b/i}$ is the measured inertial specific force of the body in the platform coordinate frame, $\tilde{\omega}^p_{b/i}$ is the platform angular velocity with respect to the inertial frame, expressed in platform coordinate frame.

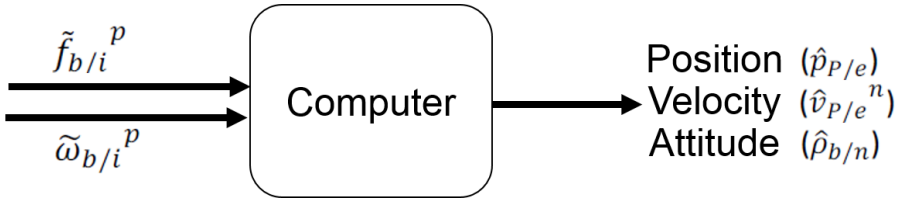


Figure 3.19: Schema of Inertial Navigation System process.

3.4.1 Principles

Inertial navigation is based on the the principle that, inside a Inertial reference frame, is possible to define the motion of a body with constant mass, observing the external forces acting on it (1^{st} Newton Law). Inertial sensors are composed of accelerometers and gyroscopes usually combined along the 3-axis in an IMU. Accelerometers measure inertial acceleration as specific force f acting on the body assimilated to a single point. This is not the actual acceleration **a** but is composed by the acceleration of the body minus the local gravity acceleration component **g**. To extract the correct acceleration value is require to know the gravity force aligned with the local vector along the vertical component at the specific time. This requires knowledge of the orientation or attitude of the accelerometers axis (along with f is measured) relative to the local g vector. This is done observing the relative rotational motion of the body with respect to the inertial reference frame using the rate of the gyroscopes. If an error is made in this rotation matrix estimation, the gravity compensation is not made properly. The attitude information is obtained from the gyros using the kinematic relationship between angular parameters (Euler angles) and angular velocity. After this, the double integration of the acceleration value in time allows to compute the position variation of the body with respect to the previous one. All these concepts are mathematically formalized by some equations of motion parametrized in the navigation frame and usually known as mechanization equations. Before to describe the mechanization process is important to define all the coordinate frames to take into account in the inertial navigation algorithms. It is possible to distinguish in:

- **Inertial frame**: the inertial frame is assumed fixed with respect the fixed stars;
- **ECEF frame**: reference frame based on an ellipsoid. The position is expressed in this frame;
- **Navigation frame**: n-frame o t-frame. Geographic frame or ned frame, is define locally and have d is normal to the plane tangent to the ellipsoid in the point. (P, n, e, d). the local reference frame have to be define time by time. The tangential frame or t-frame is defined by tangent plane to reference ellipsoid at a fixed point P_0 (P_0 is the origin of the local frame);
- **Body frame**: The body frame is attached to the vehicles (G, x_b, y_b, z_b);
- **Platform frame**: is attached with the IMU. All the measurements will be referred to the platform frame $(0_p, x_p, y_p, z_p)$;
- **Instrument frame**: it is aligned with the axes of orientation of the inner sensors itself and could be deviated from the platform frame;

- **Computing frame**: there is also a digital resolution of the pc that extract value from digital number measured by the sensors.

In order to represent a vector into different coordinate system the rotation between two coordinate system need to be computed:

$$v^a = C_b^a v^b \tag{3.68}$$

The rotation matrix from the e-frame to the n-frame is:

$$C_e^n = \begin{bmatrix} -\sin\varphi\cos\lambda & -\sin\varphi\sin\lambda & \cos\varphi \\ -\sin\lambda & \cos\lambda & 0 \\ -\cos\varphi\cos\lambda & -\cos\varphi\sin\lambda & -\sin\lambda \end{bmatrix} \tag{3.69}$$

with φ and λ geodetic latitude and longitude. The rotation matrix from the b-frame to the n-frame is:

$$C_b^n = \begin{bmatrix} \cos\psi\cos\theta & -\sin\psi\cos\phi + \cos\psi\sin\theta\sin\phi & \sin\psi\sin\phi + \cos\psi\sin\theta\cos\phi \\ \sin\psi\cos\theta & \cos\psi\cos\theta + \sin\psi\sin\theta\sin\phi & -\cos\psi\sin\phi + \sin\psi\sin\theta\cos\phi \\ -\sin\theta & \cos\theta\sin\phi & \cos\theta\cos\phi \end{bmatrix} \tag{3.70}$$

where θ, ψ and ϕ are the roll, pitch and yaw.

3.4.2 Mechanization equations

The mechanization equations are the kinematic equations representing the inertial navigation systems. These equations consist in three main steps: body position, body velocity and attitude computation:

$$\dot{p} = T\dot{v}^n \tag{3.71}$$

$$\dot{v}^n = C_b^n f^b - [(2\omega_{ie}^n + \omega_{en}^n)_\times]v^n + g^n \tag{3.72}$$

$$\dot{C}_b^n = C_b^n([(\omega_{ib}^b)_\times] - [(\omega_{in}^b)_\times]) \tag{3.73}$$

Where:

- C_b^n is the rotation of the body frame to the navigation frame;
- the **T** matrix converts the linear velocity values to angular changes in latitude and longitude:

$$T = \begin{bmatrix} \frac{1}{R_N+h} & 0 & 0 \\ 0 & \frac{1}{(R_E+h)cosL} & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{3.74}$$

  where R_N and R_E are radii of curvature of the reference ellipsoid.;

- the position vector expressed in geographic coordinates:

$$\mathbf{p} = \begin{bmatrix} L \\ \lambda \\ h \end{bmatrix}$$

- the velocity vector expressed in the navigation frame:

$$\mathbf{v} = \begin{bmatrix} v_N \\ v_E \\ v_D \end{bmatrix}$$

Implementing these equations consist in three steps: updating the attitude, updating the velocity and updating the position. Figure 3.20 illustrate the entire 3D INS mechanization procedure. More details about the implementation of the inertial mechanization are given in [55]. What is important to state is that such implementation requires an initialization of the parameters, this can be made from another sensor like GNSS. The hybridization of the INS navigation with GNSS will be presented in section 3.4.4.

Attitude computation

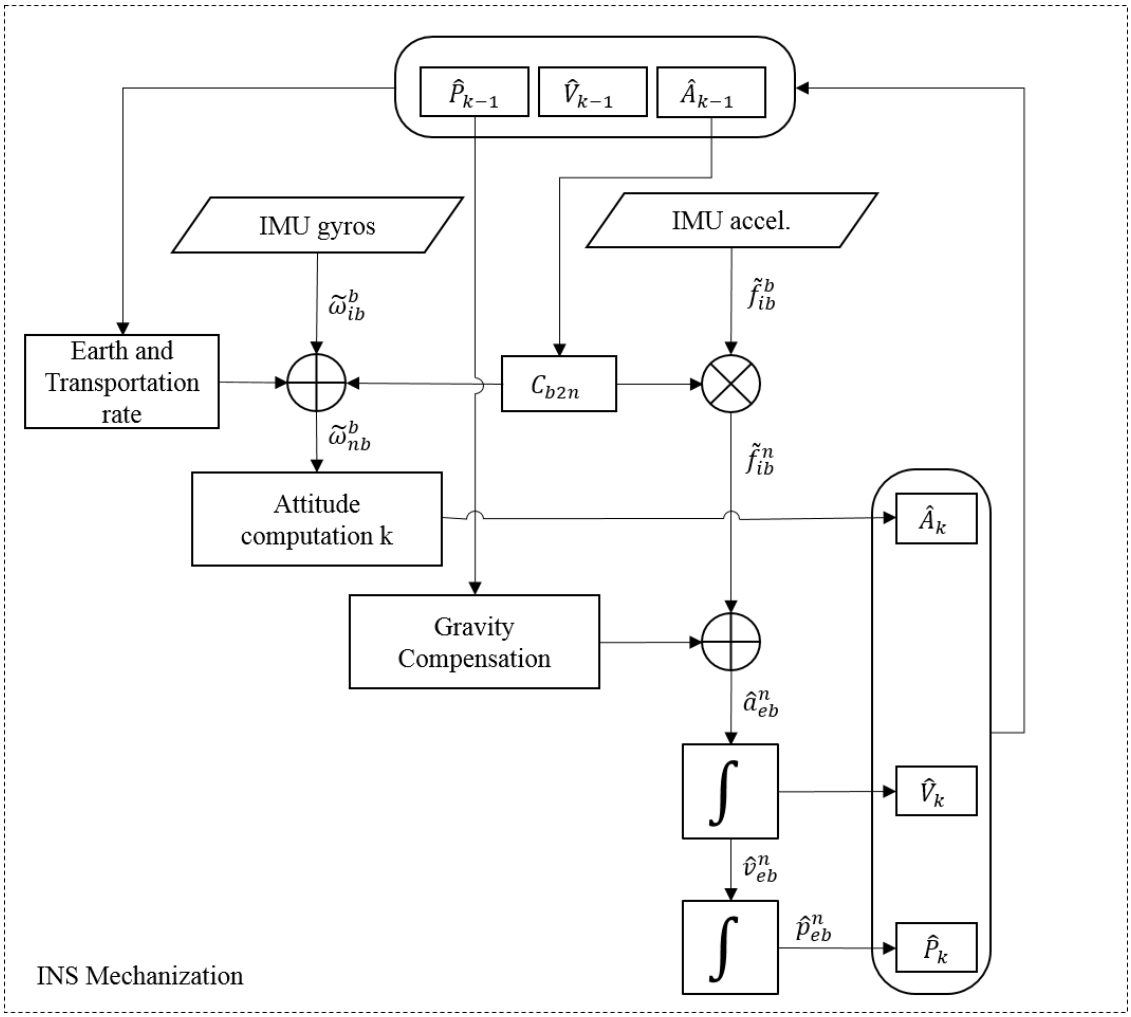


Figure 3.20: INS mechanization schema.

3.4.3 Source of INS position error

The double integration in mechanization equations makes the error following a second order polynomial function. That means that the estimation will drift very easily in few seconds. The deterministic errors affecting INS are due to manufacturing defects like misalignments, temperature dependent errors, etc. These errors can be minimized by calibration procedures [56]. The stochastic errors affecting INS are random noises on the measurements. In general, the errors in the instantaneous computed position at the INS output can arise from:

- **Instruments errors**: accelerometers and gyroscopes have noise;
- **Alignment errors**: between sensors and their platform. But also from computational process of frame transformation;

- **Environment errors**: model for the local gravity that sometimes is deviated from the true value;
- **Computational errors**: in the equation model and in the numeric integration;
- **Initialization errors**: assumed state vector;

Table 3.4 summarize the performances of different class of Inertial navigation systems. To qualify the quality of these sensor is possible to observe the magnitude of their biases. These biases are mainly constant over some times and changes only due to change of the vehicle so that the use of aiding sensors like GNSS receiver allows to estimate these biases and compensate for them.

Table 3.4: Classes of INS and relative performances.

Class	**Position performance**	**Gyro technology**	**Acc technology**	**Gyro bias**	**Acc bias**
Military Grade	1 nmi/24 h	ESG, RLG, FOG	Servo accelerometer	<0.005°/h	<30 μg
Navigation grade	1 nmi/h	RLG, FOG	Servo accelerometer, Vibrating beam	0.01°/h	50 μg
Tactical grade	< 10 nmi/h	RLG, FOG	Servo accelerometer, Vibrating beam, MEMS	1°/h	1 μg
AHRS	NA	MEMS, RLG, FOG, Coriolis	MEMS	1-10°/h	1 μg
Control System	NA	Coriolis	MEMS	10°-1000°/h	10 μg

3.4.4 GNSS/INS integration concepts

In this section the integration between GNSS technology and Inertial Navigation Systems will be presented. This integration is particularly important in the

development of this work for two reason. First, it allows to demonstrate the motivation of hybridizing sensors and fusing data for better navigation estimation with respect to stand-alone systems. Secondly its mathematical integration will be the basis for the fusion algorithm developed in this work.

Looking at the errors behaviours of INS platform and GNSS receivers is possible to observe the complementarity of the two technologies. As stated in 3.4.3, the INS output errors are time-correlated and drift very soon especially using low-cost sensors. A GNSS receiver, on the other hand has a long therm accuracy. The problem is that satellite signal is susceptible to interferences or jamming and the use of INS measurements, which are immune, can help to increase the robustness of the GNSS receiver. Another advantages of the INS is that it not requires any interaction with the environment. This is particularly useful in the situation where the GNSS can't rely on signals transmitted by the satellite constellation. This is typical in urban canyon or in indoor environment. Talking about the reference system in which the solution is provided, the GNSS gives absolute positioning in contrast with the relative estimation of the inertial platforms. In terms of data rates, INS work at higher frequency (usually more than 100 Hz) while GNSS receivers have a data rates on the order of 1-20 Hz.

In general, the integration of GNSS and inertial sensors is performed in three different hybridization architecture depending on the type of GNSS measurement used and the level of integration. These are loose, tight and ultra-tight or deep integration architectures. Figure 3.21 shows a schematic of these three integrations.

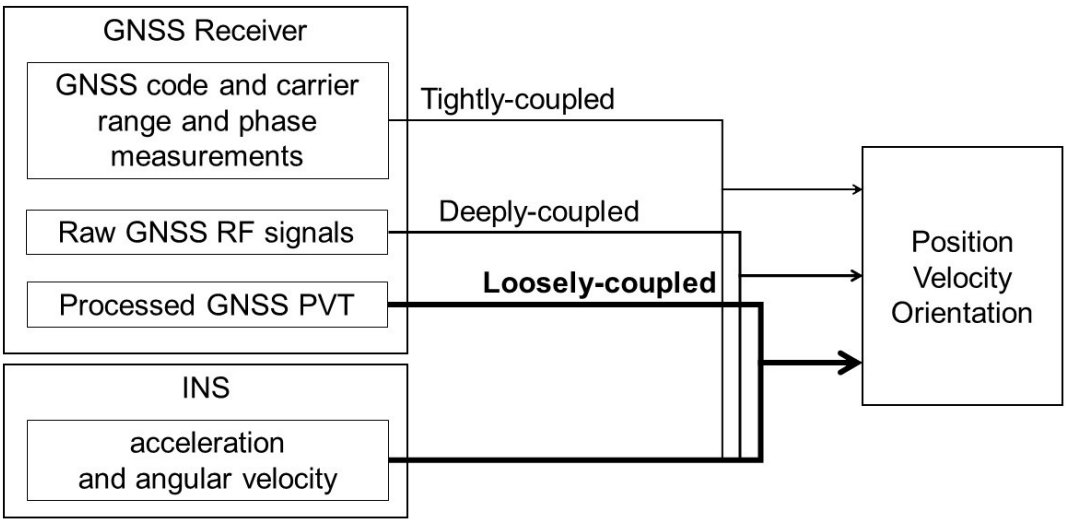


Figure 3.21: GNSS/INS integration architecture.

In the loose architecture, the INS and GNSS receivers works as independent

navigation systems. The outputs of the two systems are blended to estimate a final solution. In the tight architectures, the two sensors provides their basic measurements (pseudoranges, accelerations, angular velocity) to generate a single navigation solution. Finally in the deeply coupling the integration is made at level of the signal tracking loop. The GNSS receivers is no longer an independent sensor but INS is an integral part of the GNSS. Among the three, the loosely coupling is presented more in details in next section as it will be the basis for the navigation algorithm implementation.

3.4.5 Loosely coupling

The loose integration of GNSS and INS has been extensively investigated and consolidated during last decades and reported in several researches [57],[26],[58] and [25]. As already said, in loosely coupling integration, IMU measurements are used to compute a Position, Velocity, Attitude. This is done in a sub-system called the Inertial Navigation System (INS). The state estimate by this systems is subject by some some inertial drift or error

$$\hat{x}_{INS} = x + \delta x_{INS}$$

The GNSS measurements (position and velocity) are used in order to estimate the inertial mechanization error, within a Kalman Filter.

$$\delta \hat{x}_{INS}$$

Additionally, the IMU sensor systematic errors (bias, scale factor) are also estimated and corrected for future mechanization update.

The state vector is composed of:

- INS position error;
- INS velocity error;
- INS attitude error;
- IMU errors (bias, scale factor).

In the INS sub-system, every time one get a new measurement, the mechanization equations are used to estimate the new outputs of position velocity and attitude. As these are effected by errors which increments in time, the GNSS position and velocity are used in the filter to estimate this particular error. This correction are used to update the old state that will become new input in the INS mechanization. The measurement vector is composed of

- difference between GNSS and INS positions;
- difference between GNSS and INS velocities;

3.5 Other aiding sensors

Numerous other sensors can be integrated in the previous described methods for navigation and positioning. Some information acquired by these sensors can be aided in order to support other main technologies. The following sensors are particularly interesting in the present Thesis and their basic operating principles are reposted.

3.5.1 Magnetometers

A magnetometer sensor is able to "sense" the component of the magnetic field vector. It is a constant vector with direction and magnitude depending on the position of the body on the earth. Having the x axis of our sensor in the direction of the body movement, the magnetometer measure the component of the magnetic field vector along the x and y axis. What one want to know then is the angle between the magnetic measurement along the direction of the motion H_x and the horizontal component of the magnetic vector in the north direction H_N. To determinate this angle equation 3.75 is used. However, this is true only if the body has not any inclination with respect to the local plane, otherwise it will be affected by the residual inclination error 3.76.

$$\alpha = \arctan(H_y/H_x) \tag{3.75}$$

$$\varepsilon_{\psi mag} = \arctan(\tan(\delta)\sin(\tau)) \tag{3.76}$$

where τ is the inclination error. Other sources of errors are magnetic interferences which can deviate easily the magnetic field. These can be easily compensate with a field calibration of the sensor.

3.5.2 Baroaltimeter

Baroaltimeter, estimate the altitude as function of the pressure. The temperature decrease by knowing the physical law of the gasses and the temperature behaviour along the altitude, we can put in relation pressure measurements with altitude. The transformation from pressure to altitude is not a linear transformation but follow the International Standard Atmosphere Model (referred to P_0 at sea level):

$$\hat{h}_B = \frac{T_0}{\lambda} \cdot \left[1 - \frac{P_m}{P_0}^{\frac{\lambda R}{g}}\right], \; for \; \hat{h}_B < 11km \tag{3.77}$$

Where:

- $\hat{h}_B$ is estimated altitude above h_0,

- P_m is measured pressure
- T_0 (default 288K)
- P_0 default $101325 P_a$
- $\lambda = 0.0065°C/m$ is lapse rate or temperature gradient
- R is the universal gas constant $(287 J/(Kg.K))$
- g is accel. Of gravity $(9.80665 m/s2)$

The barometer low cost are very accurate in changing elevation. The estimated altitude is estimated above the referenced altitude.

Chapter 4

URBAN ENVIRONMENT APPLICATIONS

Positioning and navigation in urban areas suffer from different environmental issues leading to poor performances of state-of-art location systems. In harsh conditions, the quality of the navigation decrease in term of accuracy, availability, continuity and integrity until it is no longer usable for LBS. In critical domains (maritime, transport, civil aviation, military) , the main technology used for positioning is the GNSS thanks to the development of tools and models able to guarantee the requirements of the previous four criteria. These critical applications are usually performed in outdoor spaces and in no-constrained environments where the absence of obstructions and local phenomena doesn't affect the GNSS measurements. In urban cities, applications based on the location, are performed in both indoor and outdoor environments, usually navigating continuously in the passage from a closed space to an open sky area and in urban canyons i.e. outdoor areas with restricted satellite Line of Sight (LOS). Of course, when the user is located inside a building or underground, the GNSS technology is no longer available and other IPS must be used. Typical scenarios are the underground cities where numerous private and public services are located underground.

Outdoor dense populated cities presents man-made structures which affect the GNSS. In particular, the presence of trees and buildings decrease the satellite visibility inducing NLOS signal reception on the receiver (only reflected signal). Tall structures can also generate masking issues i.e. the total block of the GNSS signal. The presence of obstacles near the antenna can generate local multipath errors which degrades the quality of the direct received signal. Moreover intentionally and unintentionally radiofrequency interferences are typical of dense populated areas like urban cities.

In the interface areas i.e. in the conjunction between open and closed spaces all the issues presented above are mixed and combined so that the GNSS can be used in favourable situations while IPS must be used in other conditions. When

the objective is to perform seamless navigation none of previous methodology is able to sole the problem in stand-alone configuration but a multi-sensor integration approach must be used. It is evident that, in order to perform seamless navigation, all issues and local conditions must be taken into account for selecting the most suitable positioning methodology. This is not an easy task which requires to consider not only the positioning performances but also criteria of complexity, optimization, coverage, scalability, cost and more.

The main principle is that, in order to obtain an acceptable solution in term of accuracy, availability, integrity and continuity, a multi-sensor multi-technology approach must be considered. Due to the huge panorama of indoor and outdoor positioning technology an approach based on requirements and criteria comparison should be used to chose the bests technology given the application field. It is a multidimensional optimization problem which requires to search the best match between the parameters and the user requirements. The most important parameters are:

- accuracy and measurements uncertainty (from mm to dm level);
- coverage area i.e. the scale of the environment (single room, building, city, etc.);
- environment i.e. characteristics of the surrounding (indoor/outdoor, dark light, temperature, ect.);
- cost (cost of the sensor or total cost of the infrastructure);
- Infrastructure i.e. the required and the complexity of the sensors network (none, markers, RF sensors, GNSS);
- reference frame (local or global);
- vulnerabilities (interferences, noise, multipath, privacy, etc.);
- update rate (real time, on request, at which rate);
- power consumption ;
- number of user (single or multiple users);
- computing i.e. the memory usage and the CPU requirements.

Under these assumptions, a criteria should be selected in order to define which technology could be investigated for the aims of the research: the seamless positioning and navigation in urban environment. Observing the characteristic of the urban environment and the parameters overview characterizing the positioning systems (Section 2.2) results that the most limiting parameter for selecting a positioning

method is the coverage area of system functionality. Some infrastructure-based systems are more suitable for performing positioning in limited spaces of some hundreds of square meters like rooms, floors, plants. Other systems, which perform relative or absolute positioning can operates in larger structures like public offices, hospitals, airports and train stations. Finally, the GNSS systems is still reliable in all those situation where the user moves in large open areas, moving between interest points and in the city. Therefore, is possible to define the operational scale as the elected criteria. It has been divided in:

1. floor scale;
2. infrastructure scale;
3. district scale.

For each scale of application a positioning and navigation technology has been tested in real scenarios in order to validate the proposed solutions and to enhance limits and advantages. In particular, from the overview on related works and from the analysis on the performances of different positioning systems, the three selected technologies are UWB positioning for floor scale, Image-based localization for infrastructure scale and GNSS positioning with smartphones for infrastructure scale (Figure 4.1).

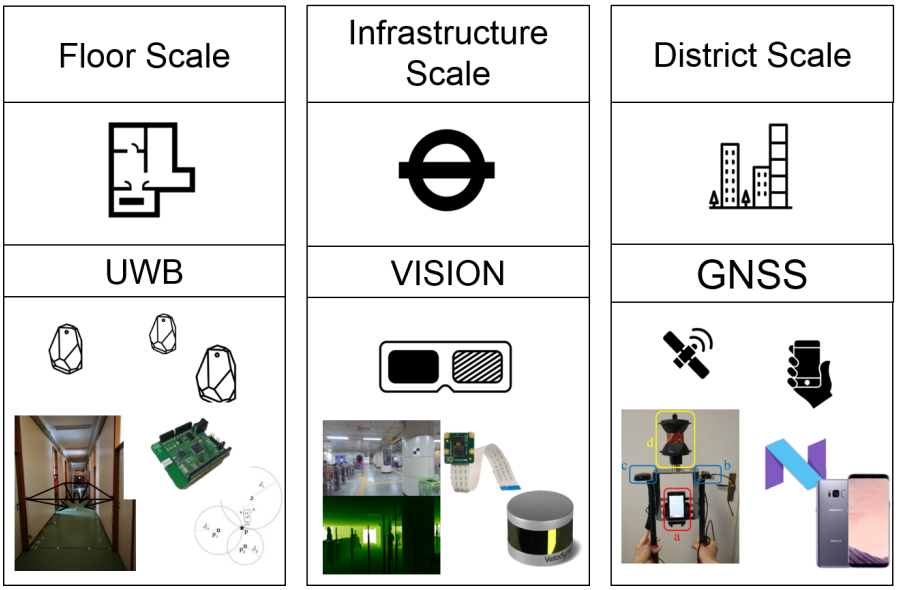


Figure 4.1: The three applications scale of urban positioning and navigation with relative selected technologies.

In this chapter three positioning methodology for urban environment applications will be presented. Regarding the "Floor Scale" an application of Ultra-wideband Indoor positioning and tracking will be presented. The reference environment is a typical office floor, with office rooms, narrow corridors, stairs. In this kind of man-made space, there are numerous interferences sources like people, furnitures or electronic devices.

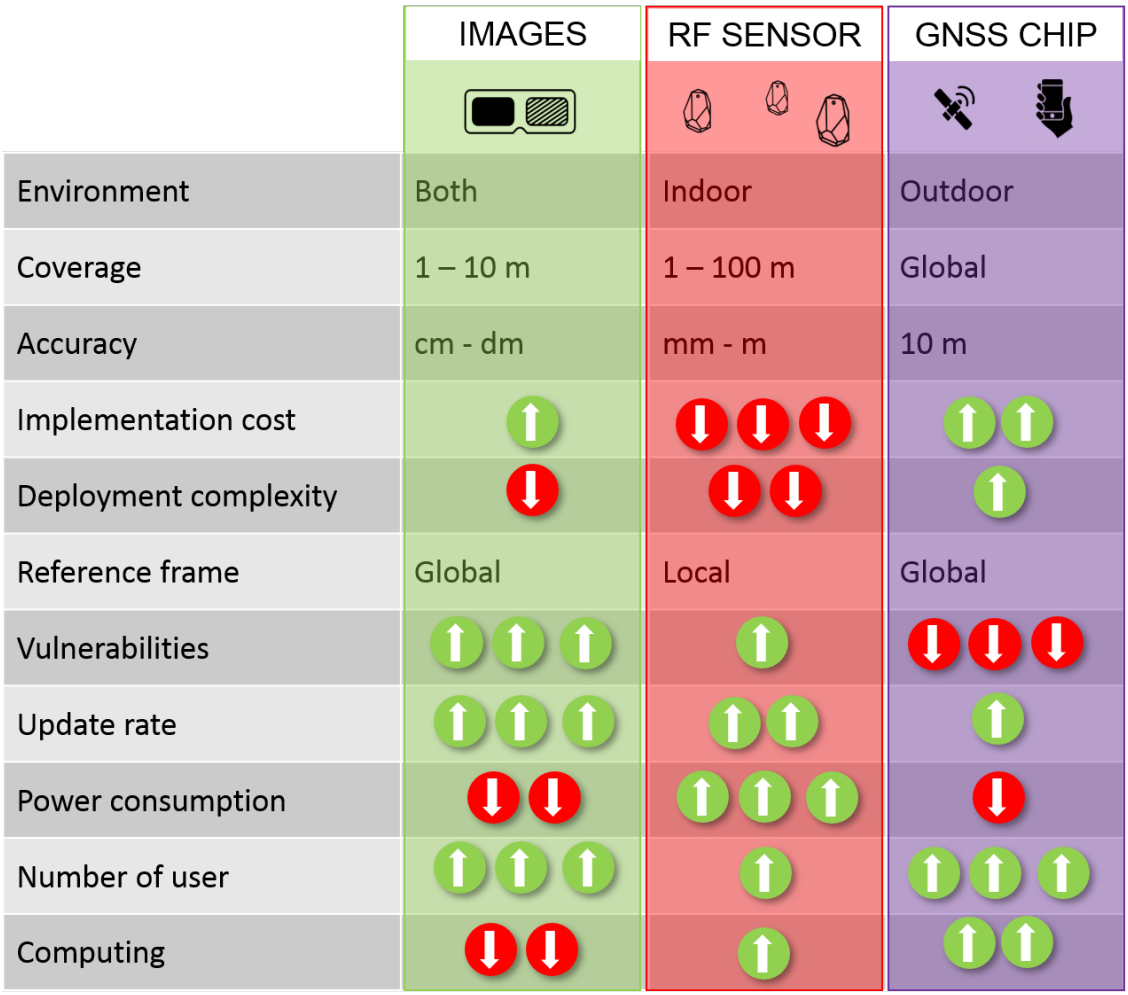


Figure 4.2: Comparative analysis on requirements parameters to performing positioning and navigation. Colors denotes advantages (green) and disadvanatges (red). Number of markers gives a comparative scale.

In the "Infrastructure Scale" an application of indoor positioning based on Imaging sensors and photogrammetric procedure is presented. Here a metro-station of a huge city and a big office building have been used as test sites. The usability of images for positioning in such crowd and repetitive environment has been investigated.

Finally, for the "District Scale" the GNSS chipsets installed in new-generation smartphones has been investigate in a urban-canyon like environment. The possibility to directly access to raw measurements has allowed to implement and perform geodetic algorithms. What is important to highlight is that these three technologies are complementary in term of coverage area and so are perfect to be integrated. Figure 4.2 shows the resulting comparative analysis made on these three systems.

4.1 UWB positioning in floor scale environment

The advantages of using UWB positioning techniques in indoor spaces are related mainly to the signal characteristics which gives to UWB-IR signal some important advantages both in telecommunication and positioning field. Firstly, as seen in Chapter 3 the UWB-IR noise-like signal makes the communication resistant to interception and detection. Then due to the very short duration pulses, the UWB signal allows to mitigate multipath reflections. The huge bandwidth allocated permits to achieve high data rates communications. Moreover, due to the low-power spectral density assigned by law, UWB signals cause very little interference with existing narrow-band radio systems like WiFi and Bluetooth and can be emitted with very low power consumption. All these advantages are reflected in the ease implementation and in the low cost of these systems. The very narrow time domain pulses mean that UWB are potentially able to offer timing precision much better than other radio systems. And this very good time domain resolution allows location and tracking applications.

In literature, there are several example of UWB technology applied in indoor environment for positioning, navigation and tracking. Ubisense, for example, developed several hardware and approaches to enhance the performances of UWB systems in different application fields. Real-time vehicle tracking [59], tracking for personnel safety in industry [60], LBS for parking [61] and more. Personnel localization for coal mines is presented in [62] while object tracking system on hospital environment is implemented in [63].

Innovative algorithms and new methodologies are investigated to improve UWB positioning performances and solve limitations like NLOS, scalability, cost and power consumption. In [64], the authors compares the performance of UWB-IR tracking systems using different positioning algorithms like weighted least square, trilateration, particle filter and Kalman filter. An extended Kalman filter approach is used in [65] to improve the navigation accuracy in highly complex indoor scenario.

More recently, the emerging interest in ITS services has increased the R&D effort on cooperative positioning. Vehicles navigation in urban environment suffer from GNSS signal interference, blockage and multipath, shortcomings that must be compensate. Extra observations transmitted from other vehicles via UWB collaborates

to improve the positioning performances. In [66], authors demonstrated a peer-to-peer cooperative localisation method able to achieve decimetre-level accuracy, by combining GNSS and terrestrial ranging measurements. A cooperative positioning algorithm to improve localisation accuracy of vehicles is presented in [67] while a distributed location estimate algorithm has also been proposed in [68]to improve the positioning accuracy by extracting inter-vehicle distance from GPS pseudo-range measurements.

In this research, low cost UWB positioning systems has been analysed and compared to evaluate the positioning and navigation performances in the floor scale previous defined.

4.1.1 UWB hardware

Numerous commercial solutions and research hardware based on UWB have been developed in last years (TimeDomain, DecaWave [69], BeSpoon[70], Ubisense location system [71], UWB-IR System of Zebra Enterprise Solutions [72]). During the present research, three different UWB sensors have been tested in several condition and compared in specific applications. Although is not the aims of this work to describe the performance characteristics and compare different commercial solutions, some performances test has been made to define the statistical parameters useful for the tuning of positioning estimation algorithms. These sensors are:

- Pozyx accurate positioning system;
- Time Domain PulsON OEM system;
- DecaWave TREK 1000 Evaluation Kit.

4.1.1.1 Pozyx System

Pozyx accurate positioning is an UWB-based hardware solution able to provide position and motion information (Figure 4.3). The system consist in a network of radiofrequency E/R modules with a very low power consumption. The 500MHz bandwidth used by this system permit range measurements with an accuracy of about 30 cm. The network is composed by one TAG transmitting the package of data, and a series of anchors with well-known position. Another TAG can be added to this configuration in order to set a master-rover configuration for the emitter module. This system transmit with a power spectrum density below -41.3dBm/MHz a train of pulse that, once accumulated, permit to the signal received to rise above the noise level. The maximum update rate for a single TAG is 80Hz (with a single TAG configuration) or around 40Hz (with base-rover configuration). Adding more TAGs decrease theis time.

Pozyx TAGs are also equipped with an IMU composed by three axis accelerometer, gyroscopes and magnetometers. With these sensors is possible to obtain the

orientation of the device. The TAGs can be connected to any computational external device, like a Raspberry Pi or Arduino boards, permitting to interact with the microcontroller unit (MCU) that manage the position algorithm. Figure 4.4 shows the hardware architecture of the Pozyx tag UWB module. The on-board UWB is a Decawave DW 1000. The systems can be controlled thank the provided libraries in ROS, Python and Arduino language.

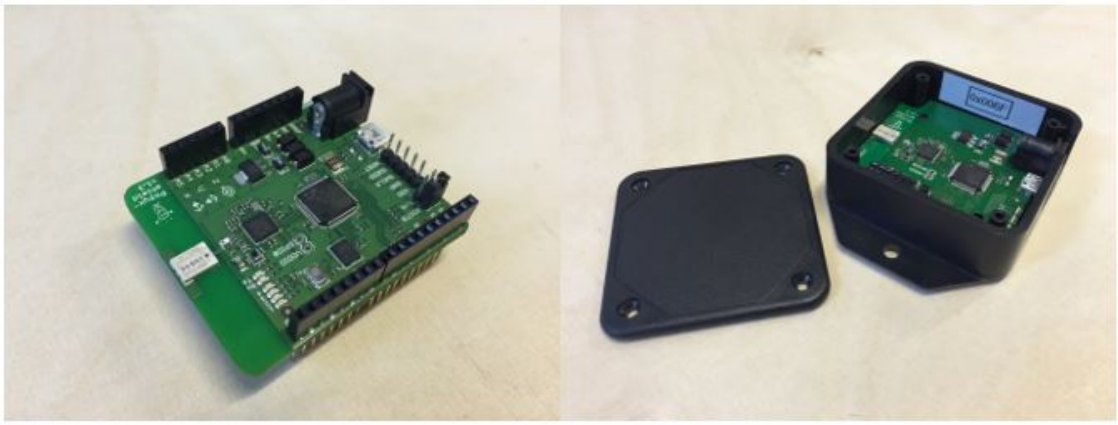


Figure 4.3: Pozyx system TAG and ANCHOR.

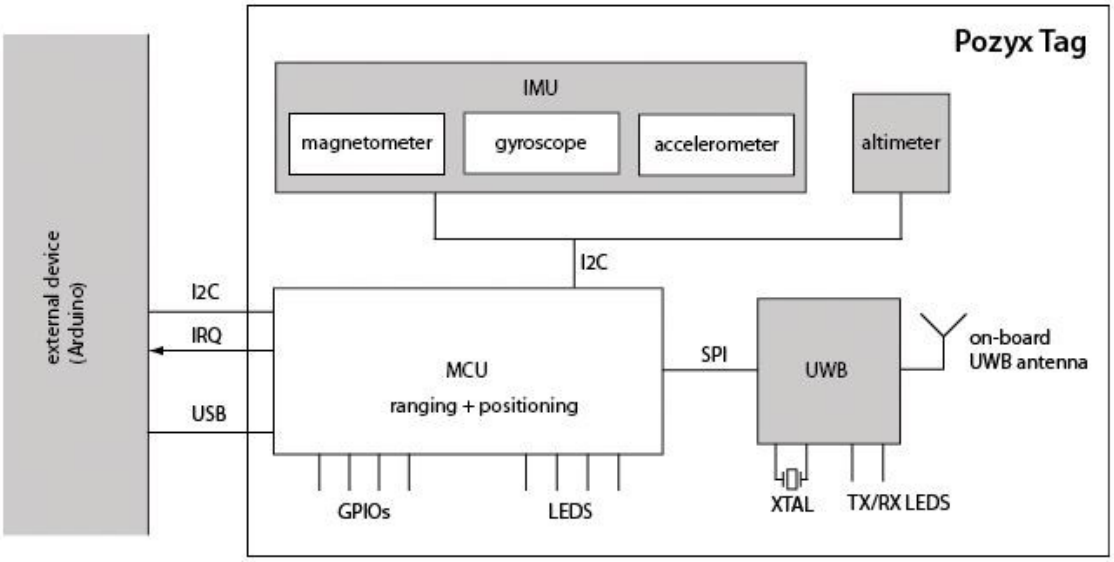


Figure 4.4: Device diagram. https://www.pozyx.io/Documentation/Datasheet.

4.1.1.2 Time Domain System

Time Domains PulsON® OEM (Original Equipment Manufacturer) system is an UWB technology hardware used to be integrated in existing product to augmented the capacity of performing ranging, tracking and positioning. It is composed by a network of UWB radio transceiver called P440 modules (Figure 4.5). It operates between 3.1 and 4.8 GHz. This system transmit with a power spectrum density below -41.3dBm/m and with a maximum update rate for ranging measurements of 125 Hz. This network operates using either the ALOHA (randomized) or TDMA (Time Division Multiple Access) protocols. Moreover the system can also operate as radar. In the UWB components there are also two connectors for the UWB antennas (Two Broadspec Antenna are provided).

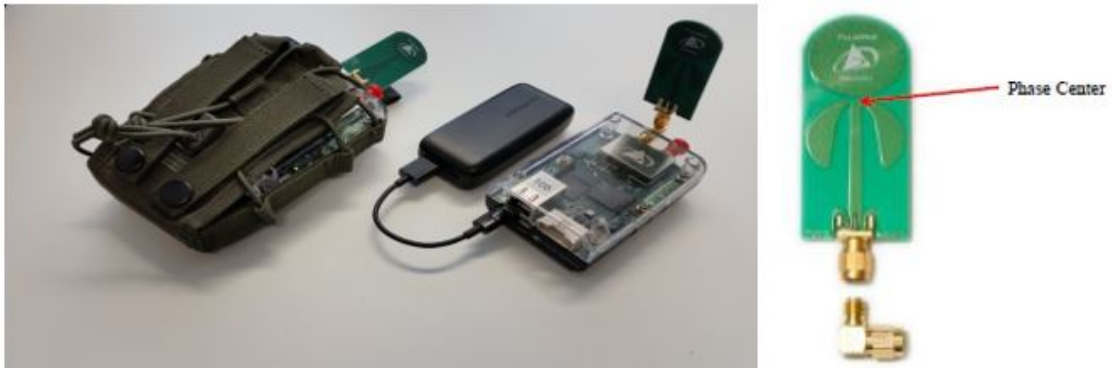


Figure 4.5: Time Domain P440 UWB module and Broadspec Antenna.

Figure 4.6 shows the hardware architecture of the P440 UWB module. It is composed by three main parts respectively the User interface, the Non-UWB components and the UWB components. The user interface part is composed by a variety of tols to physically interface every computational machine to the P440 board. Then, processor controls the UWB front end through a Digital Baseband FPGA interface. More specifically, the FPGA acts as a digital baseband to configure and control Time Domain's Fully Integrated Front End (FIFE) UWB ASIC such that it is possible to transmit and receive packets to measure range and to send/receive data.

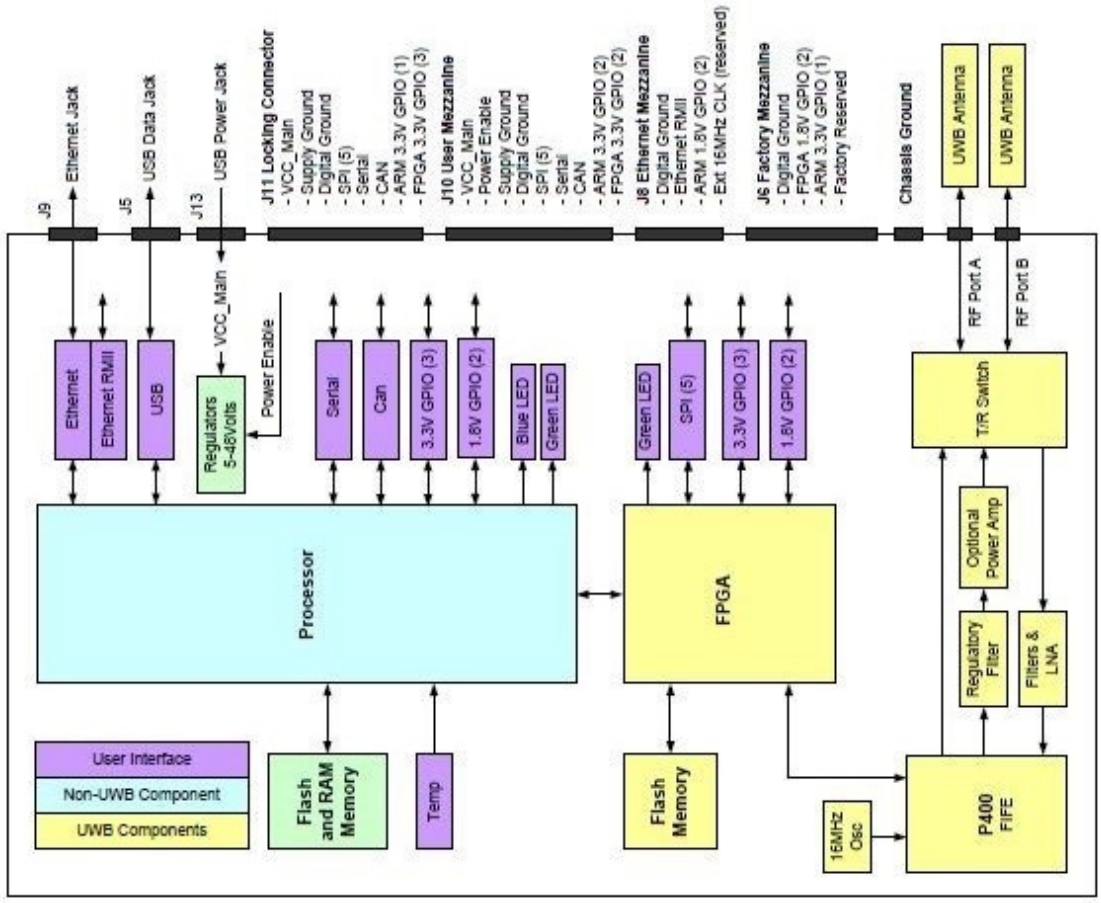


Figure 4.6: Hardware block 320-0317E P440.

4.1.1.3 DecaWave TREK 1000 Evaluation Kit

DecaWave company product used in this work is the TREK 1000 evaluation Kit, a TWR location system based on the UWB transceiver module DWM1000 (Figure 4.7). The sensor is compliant with the IEEE802.15.4-2011 technical standard. The 6 RF supported bands spans from 3.5 GHz to 6.5 GHz. The board is composed by an ARM STM Microprocessor on which a proprietary sourcecode run to perform positioning. An Application Interface is provided to control the system from any other computational unit via USB connection. Figure 4.7 shows the UWB sensors. The producer claims that accuracy of this sensor working with 1.3 GHz bandwidth is $\pm 10cm$. The maximum range is 290 meters with an update rate of 6.8 Mbit/s.

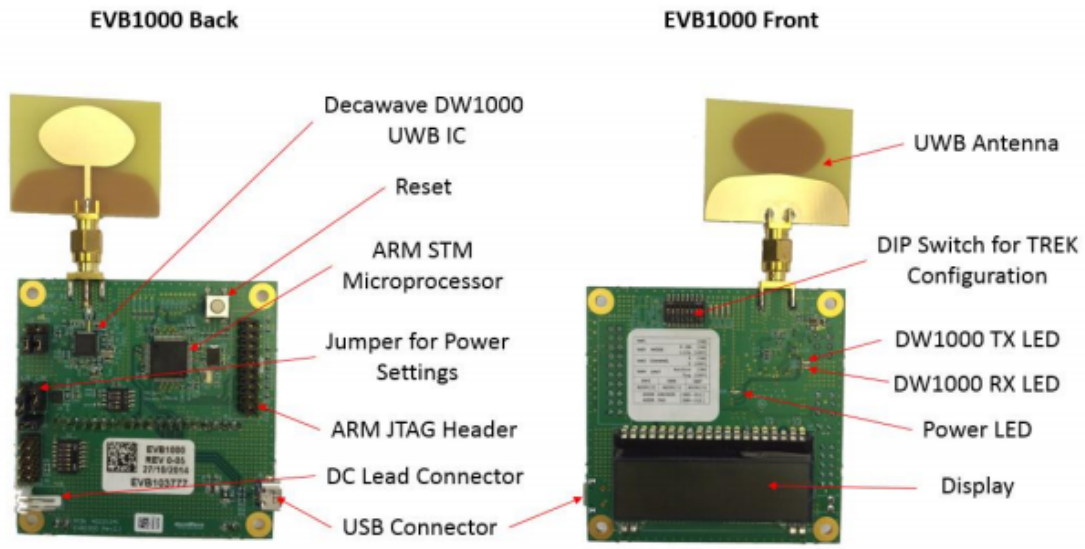


Figure 4.7: TREK1000 board details.

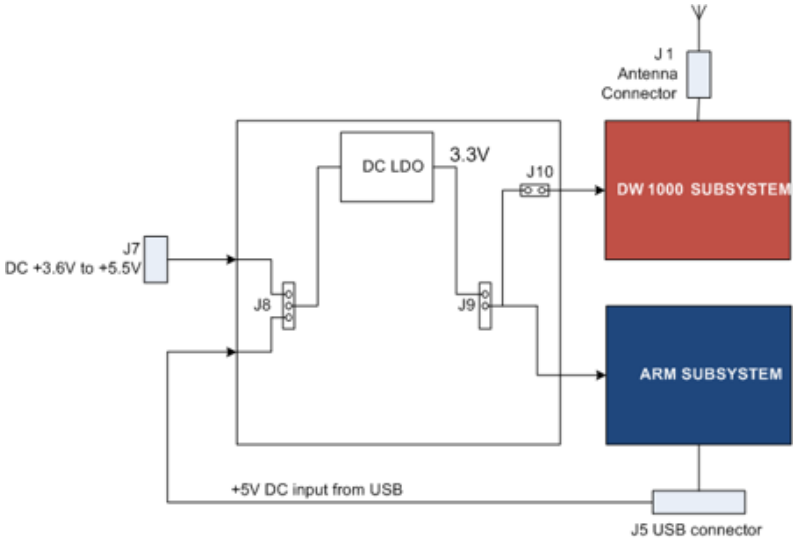


Figure 4.8: TREK1000 board schema.

4.1.1.4 Specs Comparisons

Table 4.1 compares the hardware characteristics of the three sensors presented.

Table 4.1: Hardware characteristics of UWB systems.

UWB system specs	Pozyx	PulsON P440	TREK1000
Physical Parameters			
Board Dimensions	60 x 53 mm	89 x 56 mm	70 x 70 mm
Weight	12 grams	45 grams	14 grams
Power			
Min Input Voltage	3.3 V	5.0 V with <100 mV ripple	2.8 V
Rx Power Consumption	n.a.	1.800 – 2.882 Watts	n.a.
Tx Power Consumption	n.a.	1.999 – 3.000 Watts	n.a.
RF Characteristics			
Operating Band	3.5-6.5GHz 6 Channel (500 MHz wide) Compliant with U.S. FCC mask. No certification	3.1 to 4.8 GHz 11 Channel Compliant with U.S. FCC mask. Certification has been received. Designed for compliance to ETSI 302065 mask	3.5-6.5GHz 6 Channel (500 MHz wide) Compliant with IEEE802.15.4-2011.
Center Frequency	6 channel (500 MHz wide) Ch1 : 3494.4MHz Ch2 : 3993.6MHz Ch3 : 4492.8MHz Ch4 : 3993.6MHz Ch5 : 6489.6MHz Ch7 : 6489.6MHz	4.3 GHz	min 3244 MHz - max 6999 MHz
Power spectral density	-41 dBm/MHz max	-41 dBm/MHz max	-41 dBm/MHz max
Antenna	Onboard Decawave DW1000	External Broadspec antenna	External Decawave DW1000
Ranging Performances			
Ranging Techniques	Pulsed Two-Way Time-of-Flight (TW-TOF) defined as TWR (Two-Way Ranging)	Pulsed Two-Way Time-of-Flight (TW-TOF) and Coarse Range Estimation (CRE)	Pulsed Two-Way Time-of-Flight (TW-TOF) defined as TWR (Two-Way Ranging)
Max Range	LOS: 100 m (30 m typical)	LOS: 240 m	LOS:290 m

Table 4.1: Hardware characteristics of UWB systems.

UWB system specs	Pozyx	PulsON P440	TREK1000
Precision	10 cm (typical)	Outdoors: 1 cm (typical) Indoors: 2 cm (typical)	10 cm
Bias error	2.5 cm (typical)	1 cm (typical)	n.a.
Range Meas Rate	Max 80 Hz	14 – 125 Hz	50 Hz
Localization Performances			
Localization Technique	UWB only vs Tracking (No more information provided)	Non-Linear Least Squares Fit during a short initialization period. Afterwards a Kalman Filter-based localizer incorporating a motion model and GDOP calculations or Geometric Non-Linear Least Squares Fit	n.a.
Network protocols	TDMA	n.a.	TDMA and ALOHA
Max Positioning update rate (4 anchor)	140 Hz (see table slide)	n.a.	10 Hz
Localization Accuracy (Bias & Precision):	Average horizontal error in LOS = 9.2 cm and NLOS = 14 cm	Max accuracies achieved of +/-1 mm	30 cm

4.1.2 Experimental setup

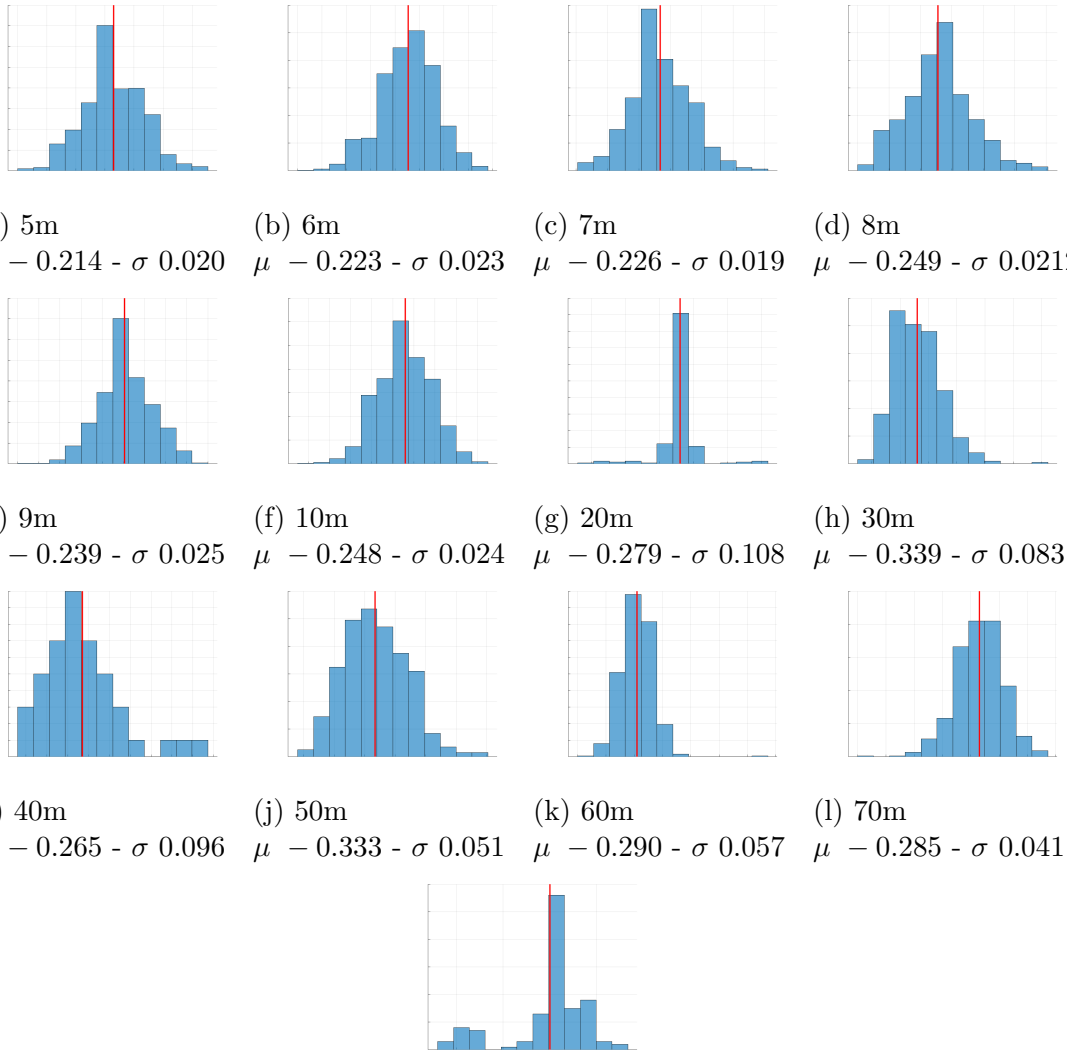
In order to evaluate the positioning capabilities of UWB technology in Room Scale environment, numerous tests have been performed on the previous presented sensors changing the environment, the network configuration, the kinematic condition and other parameters. The aim of this test is evaluate the performances in ranging measurement and positioning estimation of the UWB sensors in real case study. For Pozyx system, which implement an IMU platform, also the inertial measurements sensibility has been evaluated. All tests have been made at the Politecnico di Torino - (Italy) during several experimental campaign.

The tests can be classified in outdoor and indoor. Signal reflection has been avoided in outdoor testing while it has been stressed in crowded indoor spaces. In outdoor scenario, the maximum range measurable by a single anchor was checked together with the performances of the inertial sensors. In typical office room, the maximum positioning capability was evaluated and the sources of errors analysed. Finally in a narrow corridor, position estimation and tracking have been analysed in an unfavourable geometry configuration. In general, the algorithms provided by the sensors has been used as first check for the positioning performances. After that, more robust algorithm has been developed and tested in order to have more control on the positioning estimation.

4.1.2.1 Ranging performances

UWB sensors manufacturer declares ranging capabilities of their systems that usually don't reflects real case applications. On the Pozyx UWB system and on the TREK100 system, line-of-sight (LOS) ranging test have been performed in an open area ensuring no obstacle between two UWB sensors and sufficient distance to evaluate maximum ranging capabilities. For this test, an anchor was placed on a fixed position while a person holding the tag upwards in his hand has moved far from the anchor. The true distance between the anchor and the tag was continuously measured with a measurement tape. Fixed steps were established to acquiring some minutes of measurement in static condition. The steps were increased till reaching the maximum measurable distance which correspond to the interruption of signal communication.

From the data acquired, outlier rejection has been performed to remove recursively the measurements away from a fixed threshold. The number of samples acquired in each step ranged from a minimum of 200 to a maximum of 1200 in function of the data rates and packets loss. For each step the distribution of the measurements has been plotted as shown in Figure 4.9 and Figure 4.10.



(a) 5m
μ − 0.214 - σ 0.020

(b) 6m
μ − 0.223 - σ 0.023

(c) 7m
μ − 0.226 - σ 0.019

(d) 8m
μ − 0.249 - σ 0.0212

(e) 9m
μ − 0.239 - σ 0.025

(f) 10m
μ − 0.248 - σ 0.024

(g) 20m
μ − 0.279 - σ 0.108

(h) 30m
μ − 0.339 - σ 0.083

(i) 40m
μ − 0.265 - σ 0.096

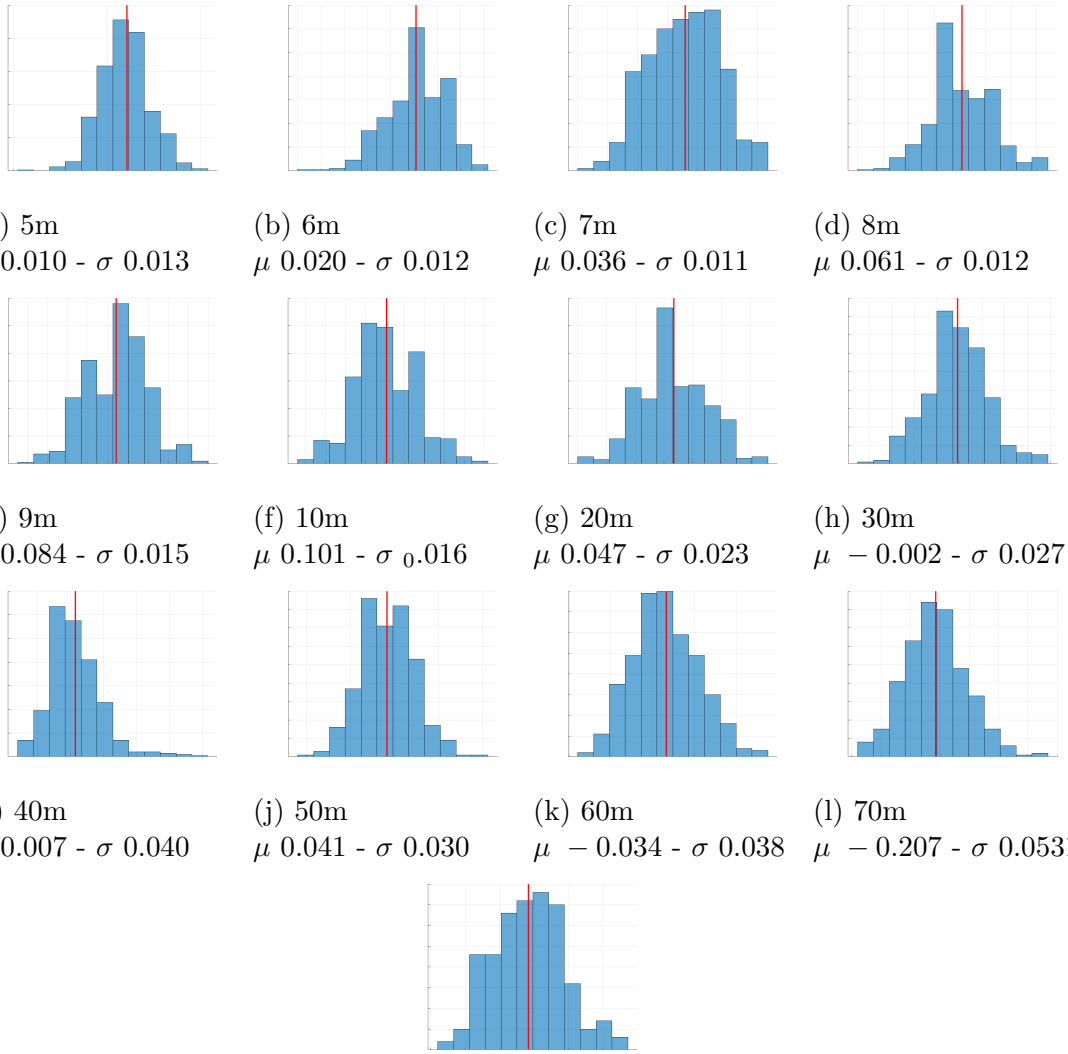
(j) 50m
μ − 0.333 - σ 0.051

(k) 60m
μ − 0.290 - σ 0.057

(l) 70m
μ − 0.285 - σ 0.041

(m) 100m
μ − 3.496 - σ 0.408

Figure 4.9: Ranging errors distibution for different reference distances expressed in meters - Pozyx UWB system.



(a) 5m
μ 0.010 - σ 0.013

(b) 6m
μ 0.020 - σ 0.012

(c) 7m
μ 0.036 - σ 0.011

(d) 8m
μ 0.061 - σ 0.012

(e) 9m
μ 0.084 - σ 0.015

(f) 10m
μ 0.101 - σ $_0$.016

(g) 20m
μ 0.047 - σ 0.023

(h) 30m
μ $-$ 0.002 - σ 0.027

(i) 40m
μ 0.007 - σ 0.040

(j) 50m
μ 0.041 - σ 0.030

(k) 60m
μ $-$ 0.034 - σ 0.038

(l) 70m
μ $-$ 0.207 - σ 0.0531

(m) 100m
μ $-$0.1083 - σ 0.0514

Figure 4.10: Ranging errors distibution for different reference distances expressed in meters - TREK1000 UWB system.

Observing these distributions is possible to state that the behaviour is mainly gaussian in contrast with no-line-of sight condition, where the error distribution is a gaussian mixture. The statistical values of the test are plotted in Figure 4.11 for the Pozyx system where is possible to observe the ranging error with respect to the real value at different distances. The results shows that both the error and the standard deviation increases with the distance, although not in a linear way. The maximum error is always less than 40 cm while the maximum standard deviation of $\pm$ 20 cm. Similar results are obtained for the TREK 1000 sensor as shown in Figure

4.13. In this case the system perform better with a mean error of maximum 20 cm (at 70 meters distance) and standard deviation always under few centimetres. For both sensor the distance from which the data communication is completely lost has been measured. The maximum operational range in LOS for Pozyx UWB system is 120 m while for TREK1000 is 157 m.

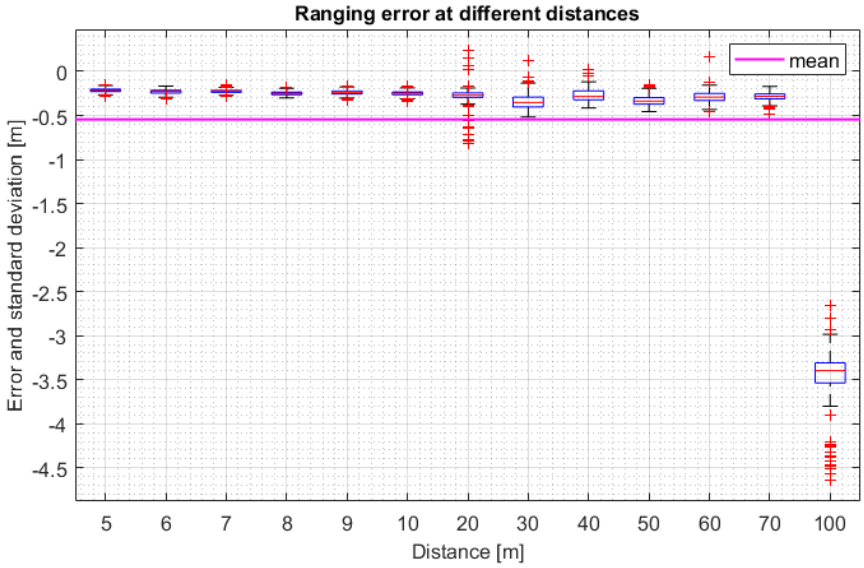


Figure 4.11: Ranging error at different distances. Pozyx UWB system.

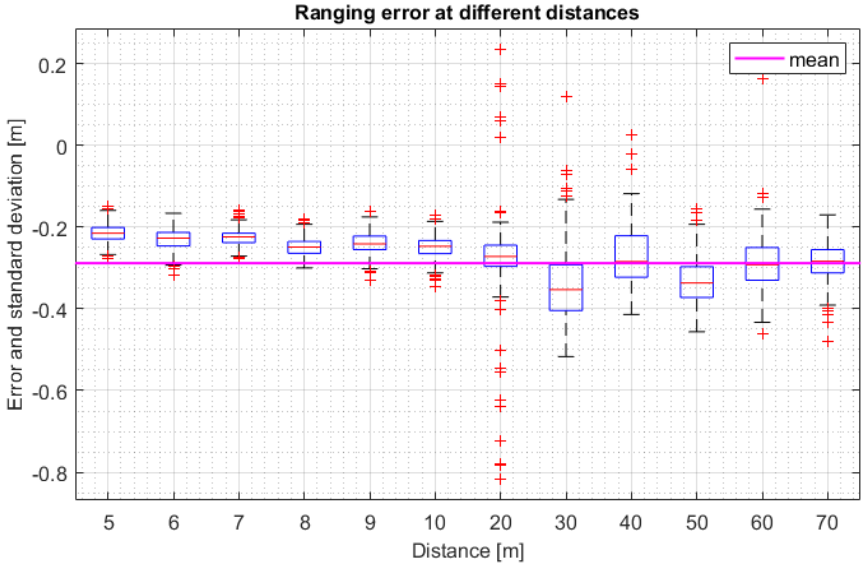


Figure 4.12: Ranging error at different distances. Pozyx UWB system.

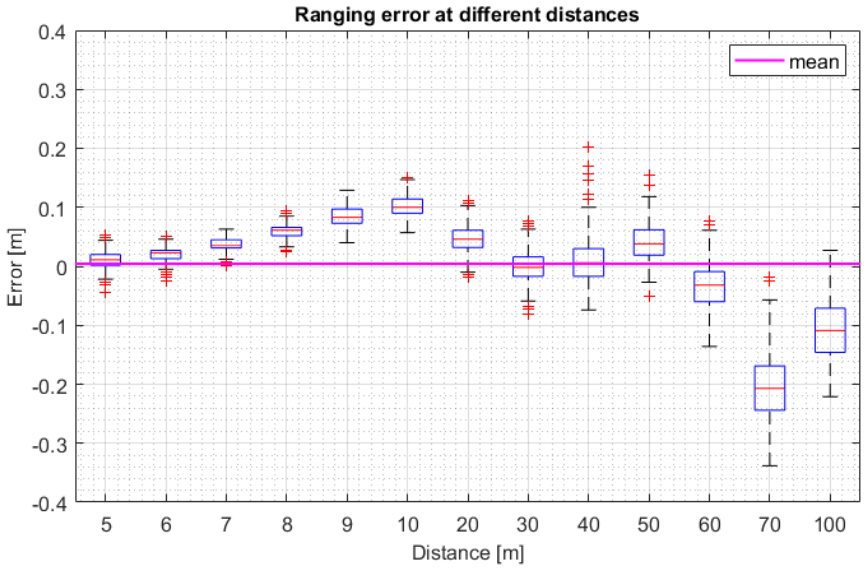


Figure 4.13: Ranging error at different distances. TREK1000 system.

4.1.2.2 Pozyx IMU performances

A comparing test on the inertial measurements has been conducted to check the internal competence of the accelerometers and gyroscopes. The comparison was made between the Pozyx system and a MicroStrain IMU (model 3DM_GX3-35). These two systems were fixed together to the telescope of a total station which is used as reference for both systems. Figure 4.14 shows the test setup. The main objective of this test was to compare orientation information acquired from both systems in terms of precision, accuracy and sensitivity.

Figure 4.14: Inertial sensors fixed on a total station.

Acquiring different raw angle measurements each 50 deg, it was observed that both systems use different scale for Roll and Heading. So, after aligning the orientation angles to a common scale, the maximum absolute difference in the orientation for Roll, Pitch and Heading is computed and compared (Table 4.2. Following, the measurements comparisons between Pozyx IMU, MicroStrain IMU and TS are presented for roll (Figure 4.15), pitch (Figure 4.16) and heading (Figure 4.17).

Table 4.2: Inertial measurements.

	Pozyx	MicroStrain
Roll	1.7 deg	0.8 deg
Pitch	4.63 deg	4.43 deg
Heading	4 deg	21.6 deg

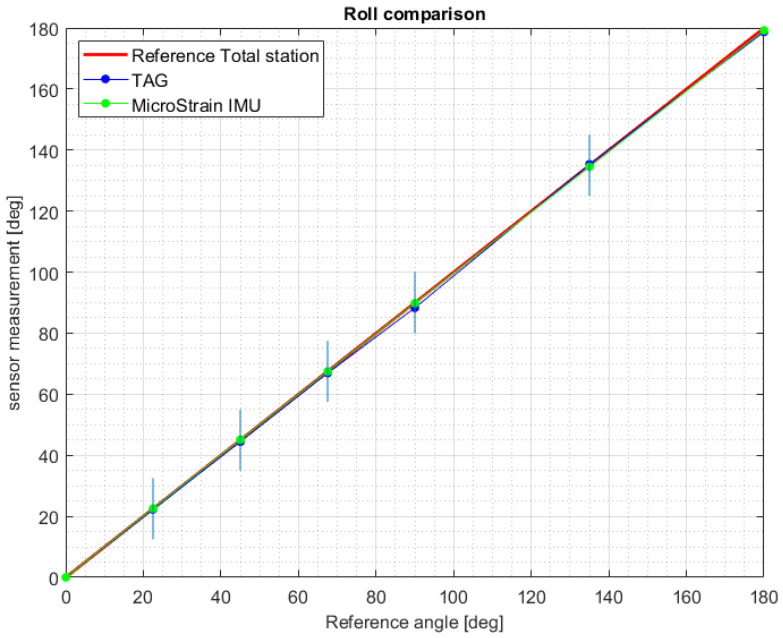


Figure 4.15: Roll estimation from inertial measurements.

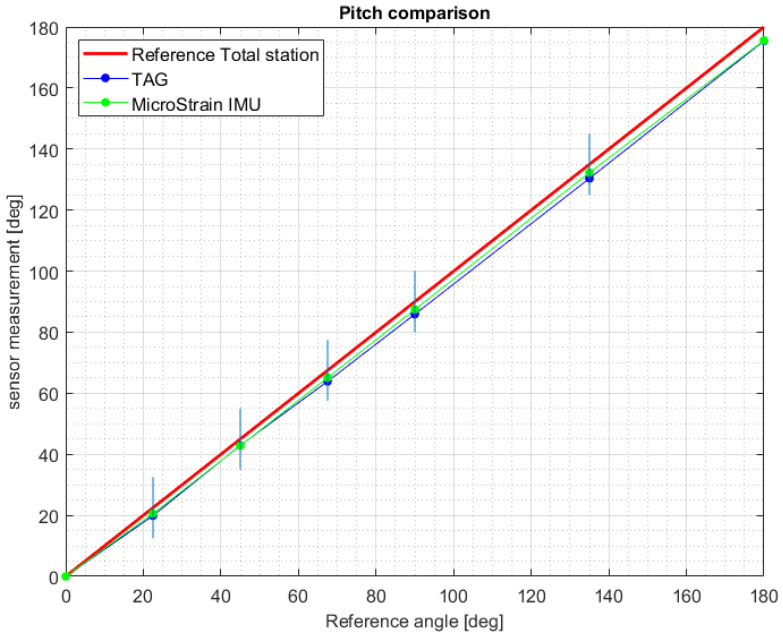


Figure 4.16: Pitch estimation from inertial measurements.

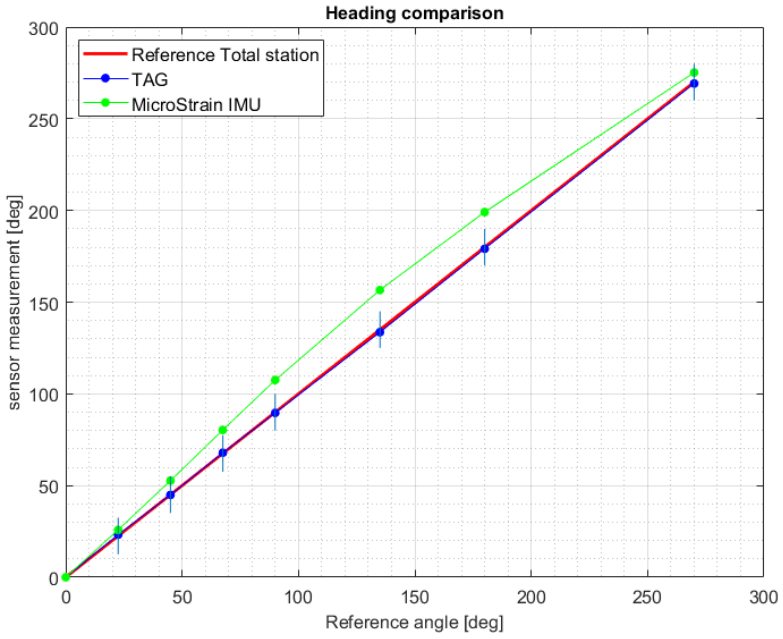


Figure 4.17: Heading estimation from inertial measurements.

4.1.2.3 Indoor Room positioning results

Testing the positioning capabilities of different UWB sensors consist in evaluate the response of the system in different environmental conditions. The parameters that influences the response of these systems are:

1. geometry configuration: Isotropic or anisotropic configuration of the UWB network changes the geometric precision as for the GNSS system;
2. number of Anchors: The number of anchors can both increase the positioning accuracy of the systems or in same cases inject noise and decrease the performances;
3. type of environment: The presence of furnitures, the passage of people, reflective or absorbing surfaces and more are environmental conditions that can affect the range measurements ans consequently the positioning behaviour;
4. motion: Also the motion is a variable to take into account as the system could perform differently in static or in Kinematic condition.

Several tests have been made changing the previous parameters and analysing the results obtained. In order to give in a glance all the tests performed Table 4.3 present the list of the experiments characteristics.

Table 4.3: List of tests performed in indoor environment. Algorithm number correspond to: 1 = UWB_ONLY, 2 = TRACKING, 3 = NLLSE, 4 = Inner.

Environment	**Indoor Office Room**										**Indoor Geomatic Lab**	
Sensor	Pozyx										Pozyx	TREK 1000
Motion	Static						Kinematic				Kinematic	
Anchor number	4			6			4		6		4	
Algorithm	1	2	3	1	2	3	1	2	1	2	1	4

In the office room, test were performed within an office space of 6.44 m x 4.91 m. Four anchors were placed with different height in the four corner of the room and this position was measured with a total station. One of the four anchor has been fixed as origin of the coordinate system $(0\,0\,Z)$. The room was selected to avoid obstacles in order to assess maximum positioning capabilities of the system. On the floor of the office several points were materialized that maintains line-of-sight (LOS) from all the anchors in order to verify the variations in precision and accuracy changing the position. Also a 6-anchors configuration was tested.

All the reference points and the anchor position has been measured with a millimeter level of precision raised from a topographic survey. The data acquired from the sensors are both positioning estimation from the inner algorithm computation and raw ranges from each anchor receiver. On these data several analysis have been

made both for static and kinematic applications. Figure 4.18 and 4.19 shows the floor plan of the testsite, the local coordinates of anchor fixed sensors and static position of the Tag. Table 4.4 summarize all the test performed in the office room.

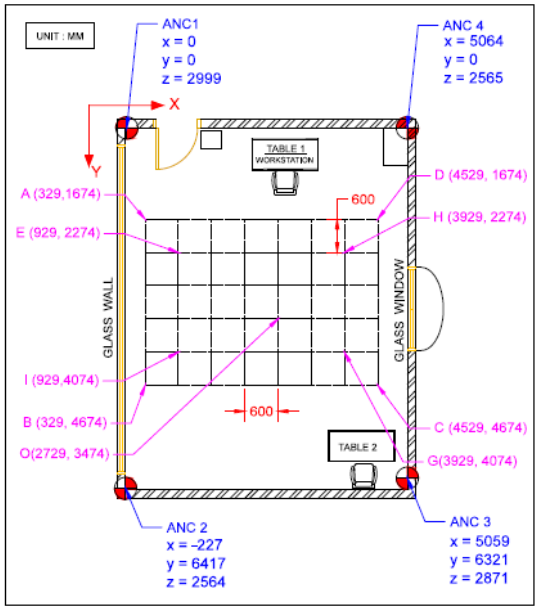


Figure 4.18: Room floorplan with 4-anchors configuration.

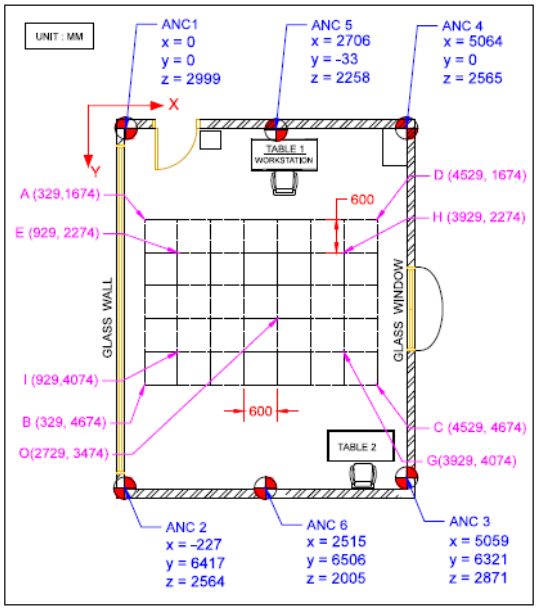


Figure 4.19: Room floorplan with 6-anchors configuration.

Table 4.4: List of tests performed in the office room.

Motion	**Static**	**Static**	**Kinem.**	**Kinem.**	**Kinem.**	**Kinem.**
N. of Anchor	4 Anchor	6 Anchor	4 Anchor	6 Anchor	4 Anchor	4 Anchor
Pos. Error	2D and 3D	2D and 3D	-	-	-	-
Pos. be-haviour	-	-	2 walking path	2 walking path	Random Walk	Outside Network

1. **Configuration 4-anchor in Static condition**

   This test was performed acquiring data for some minutes standing on eight different points spread around the room. These points, measured with high accuracy, has been used to perform statistic evaluation on the inner algorithms provided by the commercial sensor. Moreover, with the acquired raw data, a Least Mean Square estimation of the position has been performed. Figure

4.20, 4.21 and 4.22 shows the eight points true position and the estimated coordinates with the relative XY errors. Figure 4.23 and 4.24 compares the positioning error estimation both in 2D and 3D for all the points. 80 samples has been compared for the three estimation algorithm. The simmetry of the network is well described by the bias behaviour of the eight points.

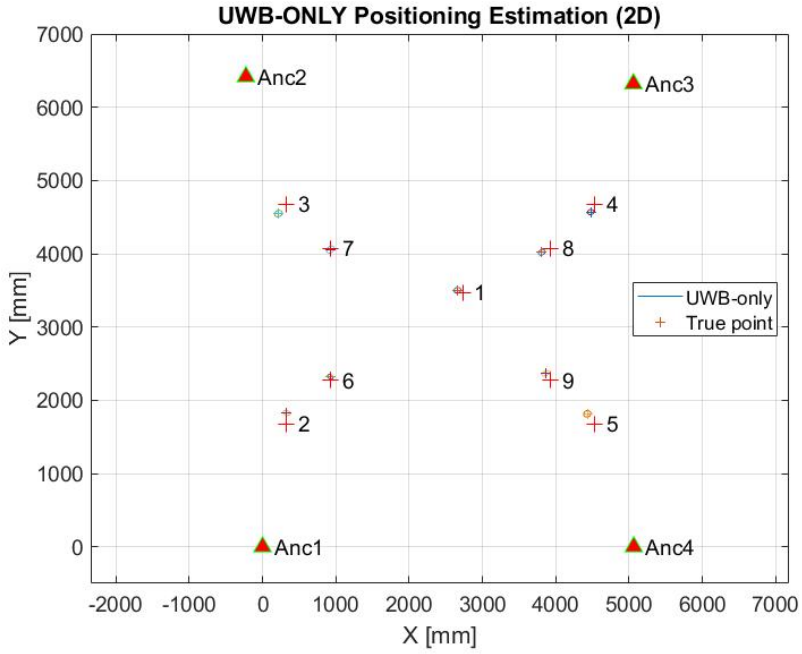


Figure 4.20: True vs estimated position of eight points for UWB-ONLY algorithm.

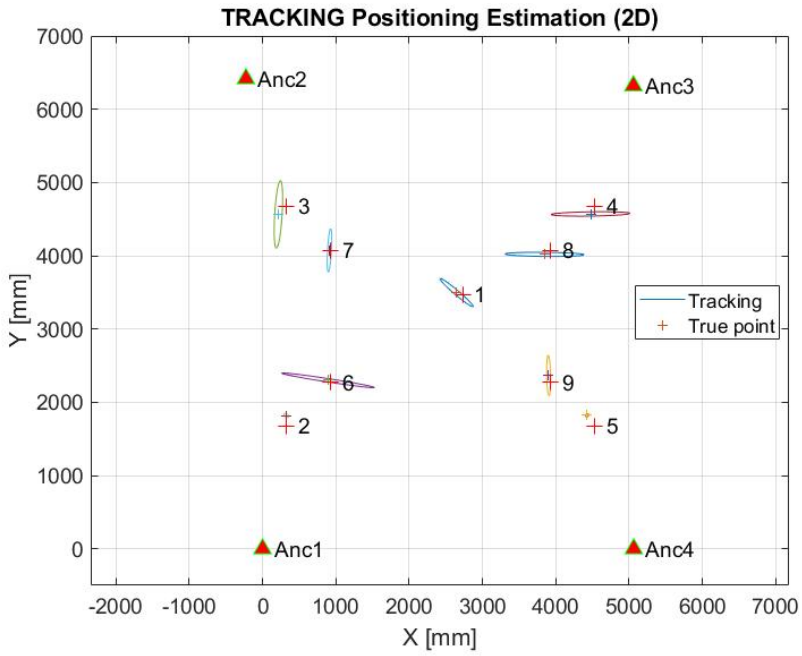


Figure 4.21: True vs estimate position of eight points for TRACKING algorithm.

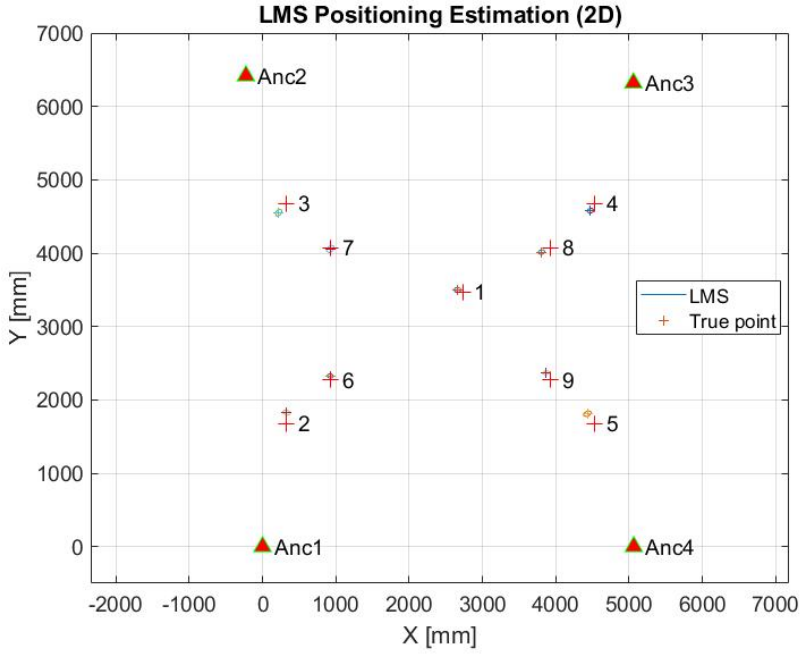


Figure 4.22: True vs estimated position of eight point for LMS algorithm.

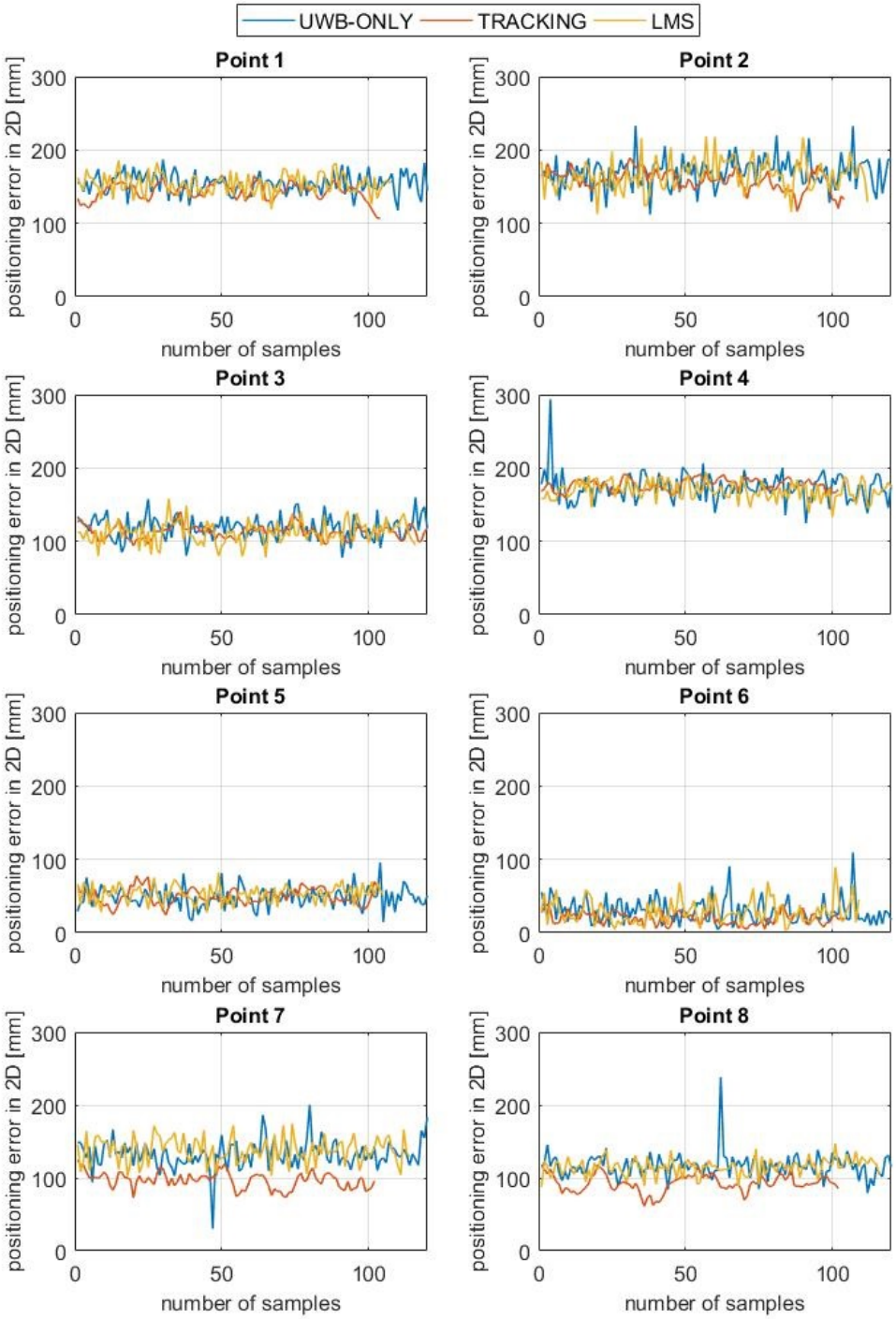


Figure 4.23: Positioning error comparison in 2D.

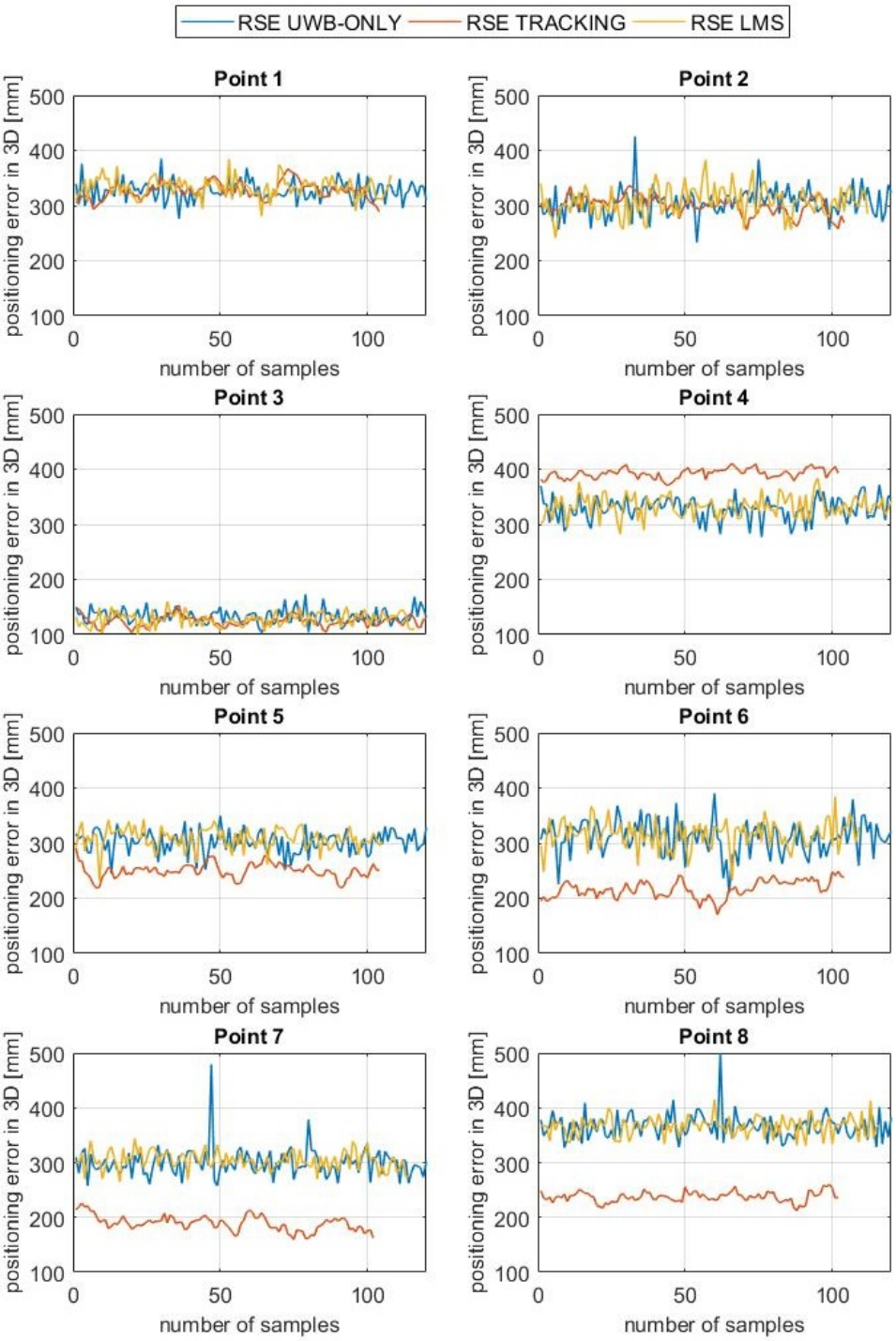


Figure 4.24: Positioning error comparison in 3D.

Table 4.5 summarize the statistical parameters of the three algorithms in 2D

while Table 4.6 compares it in three dimensions. The accuracy is interpreted as the difference between the true value of a certain point (measured with the total station) and the computed one obtained by the system. On static condition is possible to observe how the three algorithm gives almost the same results with a three-dimensional error of around 30 cm. Moreover is possible to observe how the use of the TRACKING algorithm in the static positioning estimation, inject noise on the observables so that the standard deviation increase of about 10 cm. As the localization accuracy of the considered systems are largely dependent of the number of device and their spatial configuration, the reported accuracy is intended as the typical achieved accuracy of the system.

Table 4.5: Statistic analysis on 2D positioning error.

STATISTIC	**UWB-ONLY**	**TRACKING**	**LMS**
Min [mm]	4.47	3.33	3.26
Max [mm]	293.88	2743.61	218.55
Mean [mm]	119.53	130.11	123.05
Median [mm]	127.97	110.00	133.87
St.D [mm]	52.67	159.71	50.31

Table 4.6: Statistic analysis on 3D positioning error.

STATISTIC	**UWB-ONLY**	**TRACKING**	**LMS**
Min [mm]	103.71	103.32	99.10
Max [mm]	502.18	2837.50	415.19
Mean [mm]	303.97	257.38	307.06
Median [mm]	317.48	235.36	322.24
St.D [mm]	63.66	164.38	63.98

Finally Figure 4.25 shows the CDF comparison of the three algorithms.

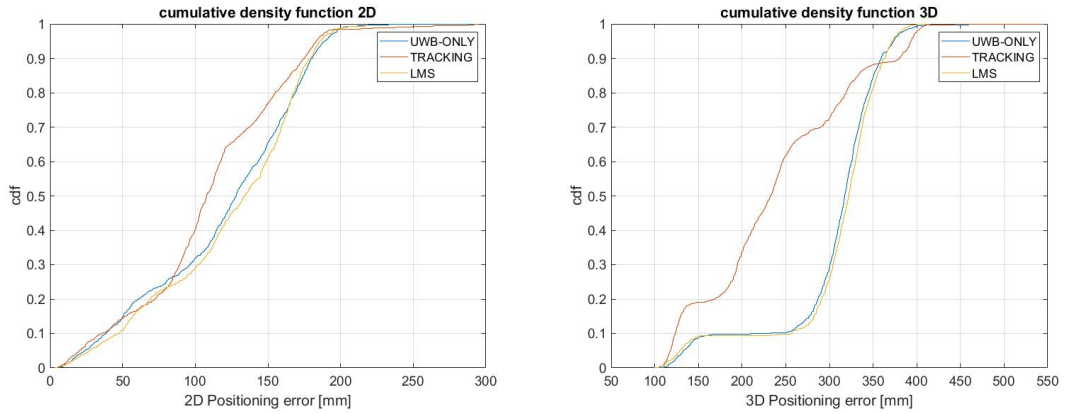


Figure 4.25: CDF of the positioning error in 2D and 3D.

Figure 4.25 shows the results obtained by the two inner algorithms provided by the sensor itself. As already said, the "TRACKING" algorithm integrate inertial measurements in the estimation of the position of the TAG. As one can expected in static condition, the figures shows the convergence of both the algorithm to the same estimation.

2. **Configuration 6-anchor in Static condition**

   The same tests have been performed upgrading the number of anchor from 4 to 6 in order to verify an increase on the positioning accuracy. The results shows that the discrepancies between the real value and the computed one for both the configurations, follow a quite random behaviour. In some of the point the accuracy is improved while in other it remains almost the same or is reduced. Keeping the same configurations and adding two more anchors does not significantly change the accuracy of the system.

   This test was performed acquiring data for some minutes standing on eight different points spread around the room. These points, measured with high accuracy, have been used to perform statistic evaluation on the inner algorithms provided by the commercial sensor. Moreover, with the acquired raw data, a Least Mean Square estimation of the position has been performed to verify the inner algorithms and to have more control on the estimation procedure. Figure 4.26, 4.27 and 4.28 shows the eight points true position and the estimated coordinates with the relative XY errors.

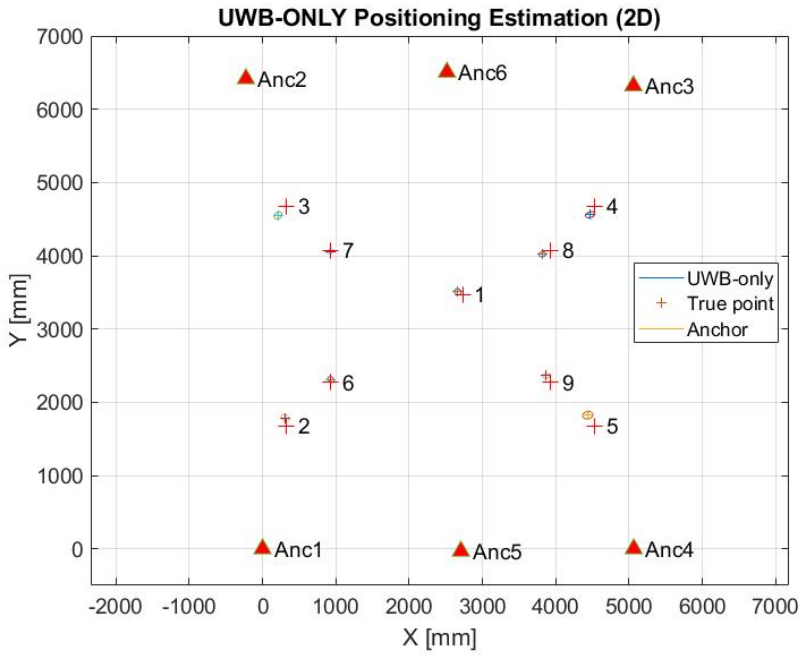


Figure 4.26: True vs estimated position of eight points for UWB-ONLY algorithm.

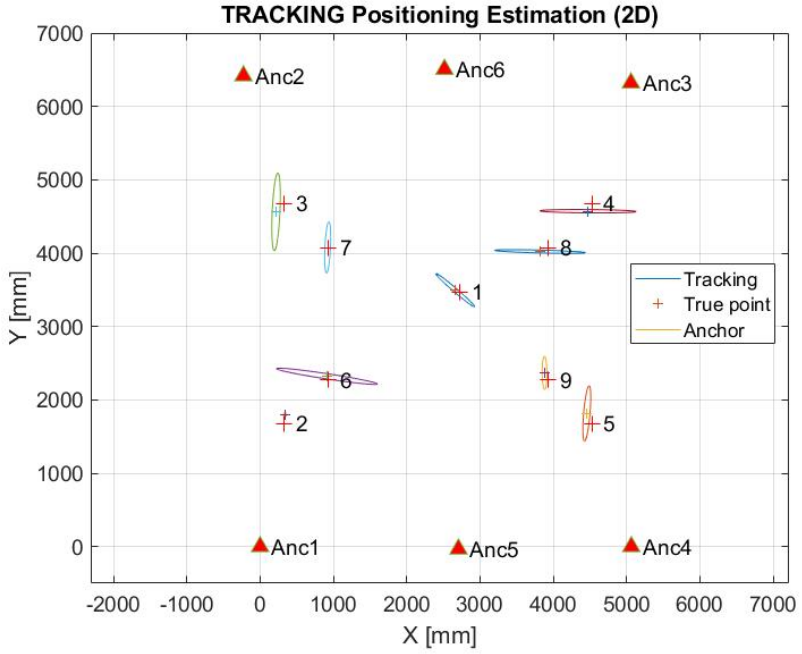


Figure 4.27: True vs estimate position of eight points for TRACKING algorithm.

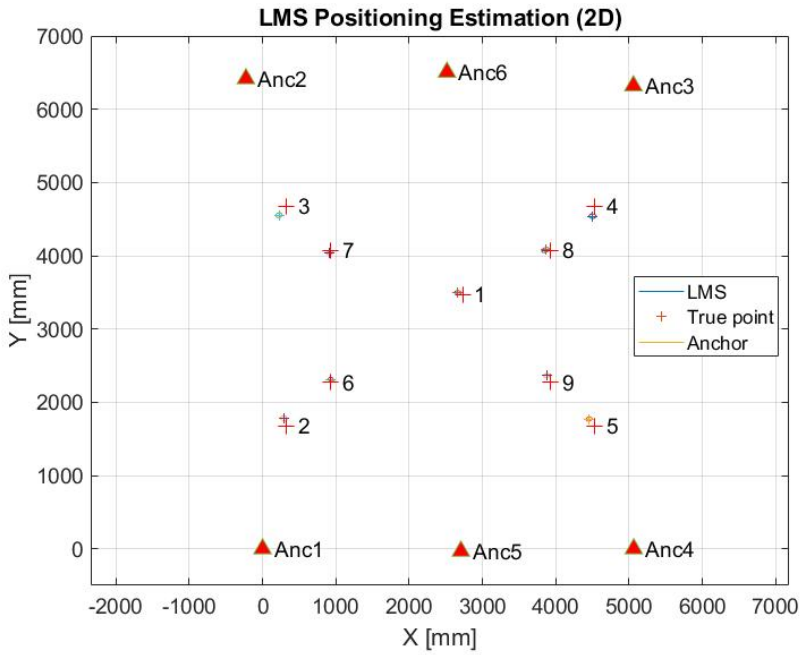


Figure 4.28: True vs estimated position of eight point for LMS algorithm.

Figure 4.29 and 4.30 compares the positioning error estimation both in 2D and 3D for all the points. Table 4.7 summarize the statistical parameters of the three algorithms in 2D while Table 4.8 compares it in three dimensions.

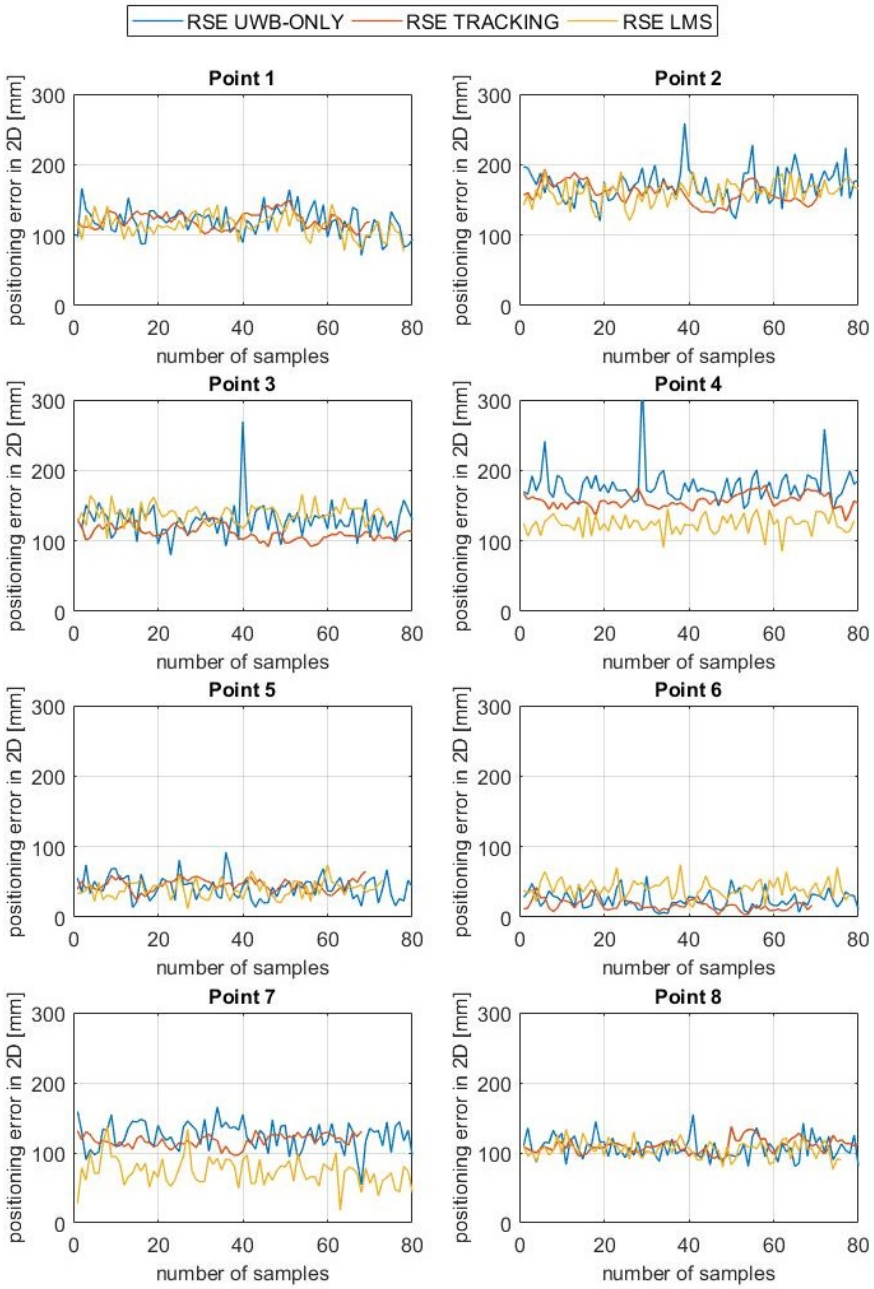


Figure 4.29: Positioning error comparison in 2D.

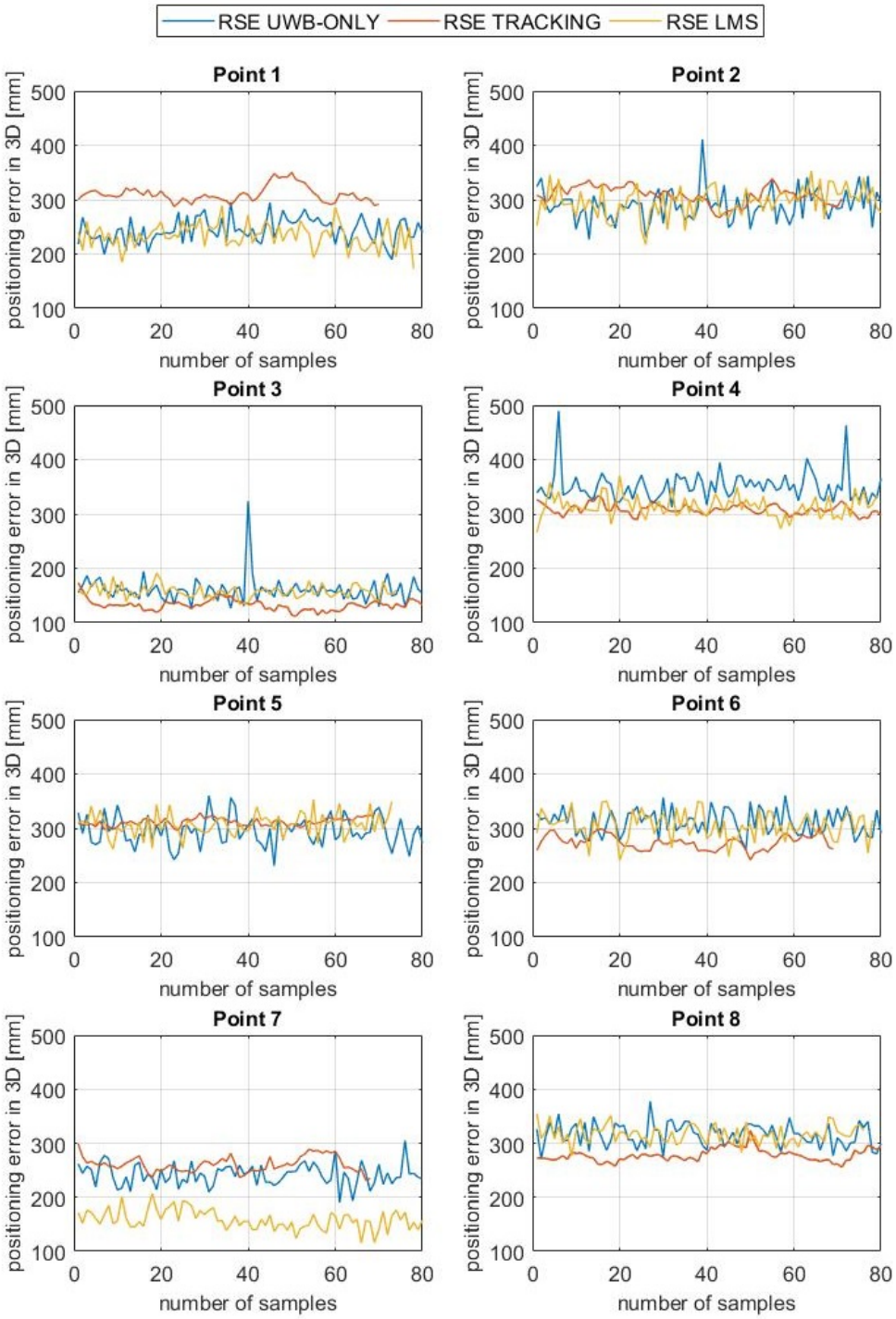


Figure 4.30: Positioning error comparison in 3D.

Table 4.7: Overall 2D statistic results.

Statistic	UWB ONLY	TRACKING	LMS
Min [mm]	2.49	3.33	11.73
Max [mm]	324.68	2860.19	192.33
Mean [mm]	107.71	137.64	102.77
Median [mm]	113.90	114.77	108.14
St.D [mm]	53.60	176.87	45.16

Table 4.8: Overall 3D statistic results.

Statistic	UWB-ONLY 3D	TRACKING_3D	LMS_3D
Min [mm]	126.46	111.821	115.68
Max [mm]	489.70	2955.35	370.87
Mean [mm]	278.74	292.44	265.69
Median [mm]	288.5318	295.66	288.72
St.D [mm]	57.96	170.19	62.10

Figure 4.31 shows CDF of the three algorithms.

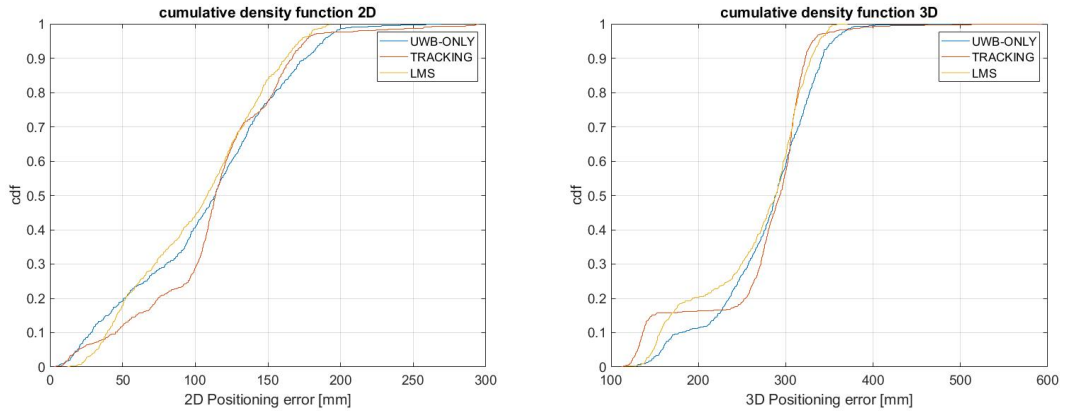


Figure 4.31: CDF of the positioning error in 2D and 3D.

3. **Configuration 4 and 6-anchor in Kinematic condition**

   In the same environment, a kinematic test was performed. The rover is moved on the marked points to assess the system for kinematic positioning. Two different path has been follow for both "UWB_ONLY" and "TRACKING" algorithm. Again, the test has been made both for 4 and 6-anchor configuration.

Figure 4.32 shows the results in 2D positioning. Along the line AB it is possible to observe the deviation from the true track which could be due to the presence of a thick glass wall along this path that causes reflections thus reduces the accuracy. Moreover, in kinematic condition, the integration of the inertial measurements allows a smoothed path estimation.

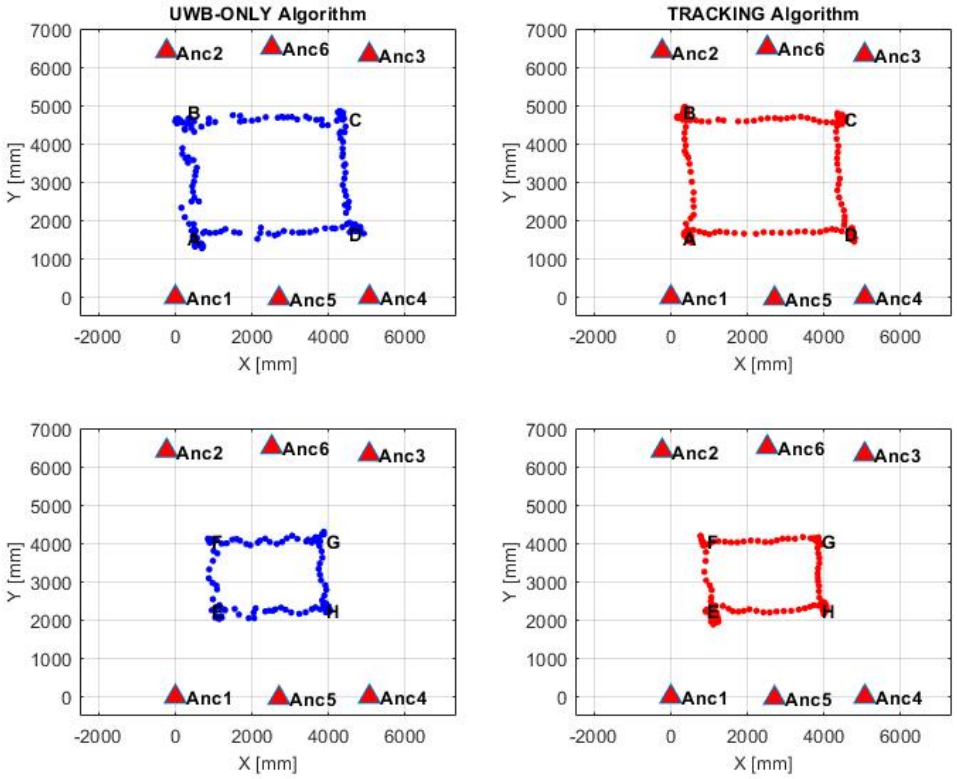


Figure 4.32: Positioning estimation with 6 anchor configuration in kinematic mode.

4. **Kinematic condition walking outside the network**

   Finally, the positioning performances of the system when the tag goes outside from the network has been observed and reported in Figure 4.33. Together with the increase of the errors in the 2D positioning for both algorithms, is also interested to observe the variation on the vertical axis. Both solution diverges from the ground truth but with an opposite sign. While "UWB_ONLY" algorithm estimate a Z coordinates minor than the real one, the "TRACKING" algorithm increase these estimates. Figure 4.34 shows the results in 3D visualization.

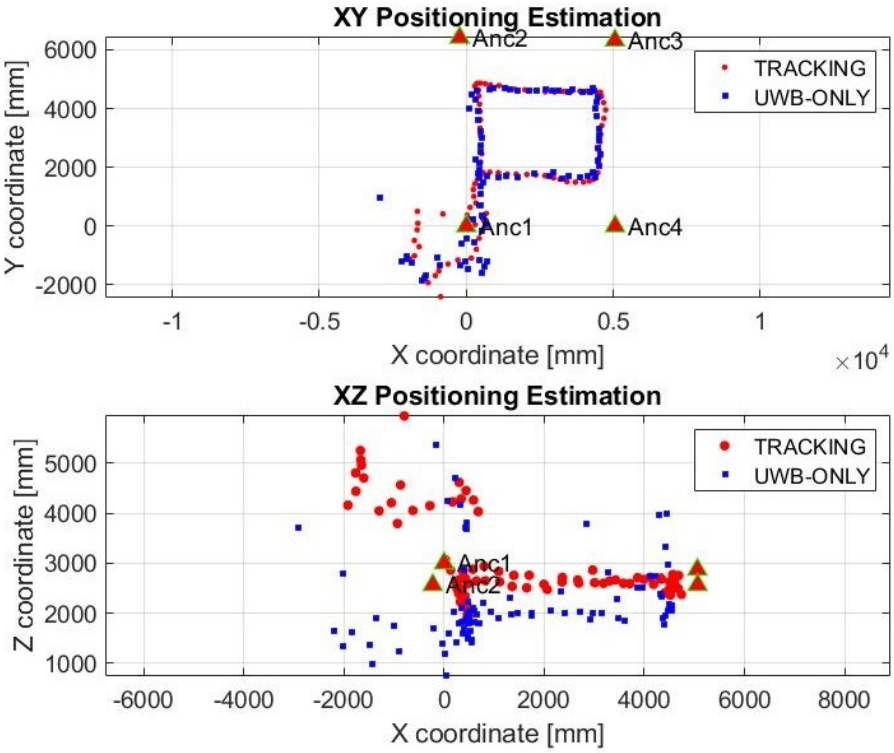


Figure 4.33: Moving outside the network of anchors.

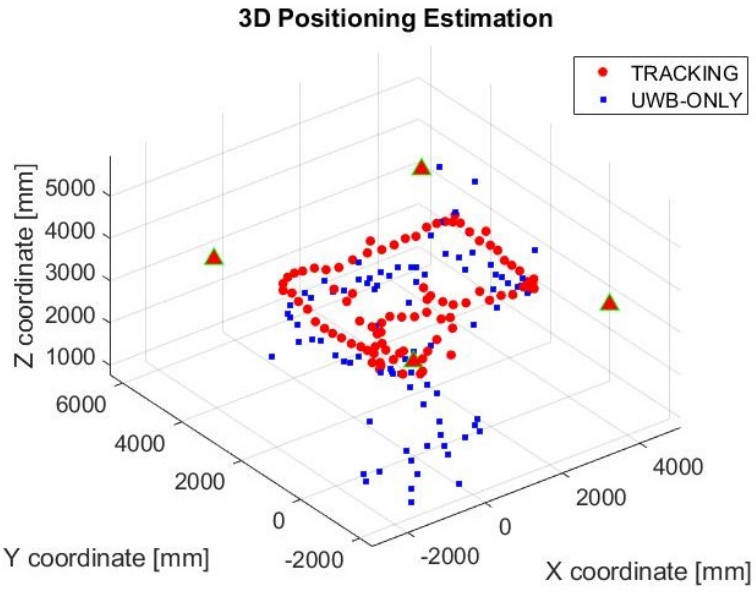


Figure 4.34: Moving outside the network of anchor.

4.1.2.4 Indoor narrow corridor positioning results

In this test, system is examined in a narrow corridor in the presence of numerous interferences such as glass windows, metallic rail, power socket and narrowness of the passage to check its competence in a harsh environment. In this site, 4 anchors are installed in a short network of 1.8 m x 6.8 m dimension (Figure 4.35). Five different points are marked in the test field for positioning. This was a static test on which the UWB_ONLY algorithm provided by the Pozyx system has been tested. After that, also a non linear least square approach has been performed to obtain the position estimation on the five points.

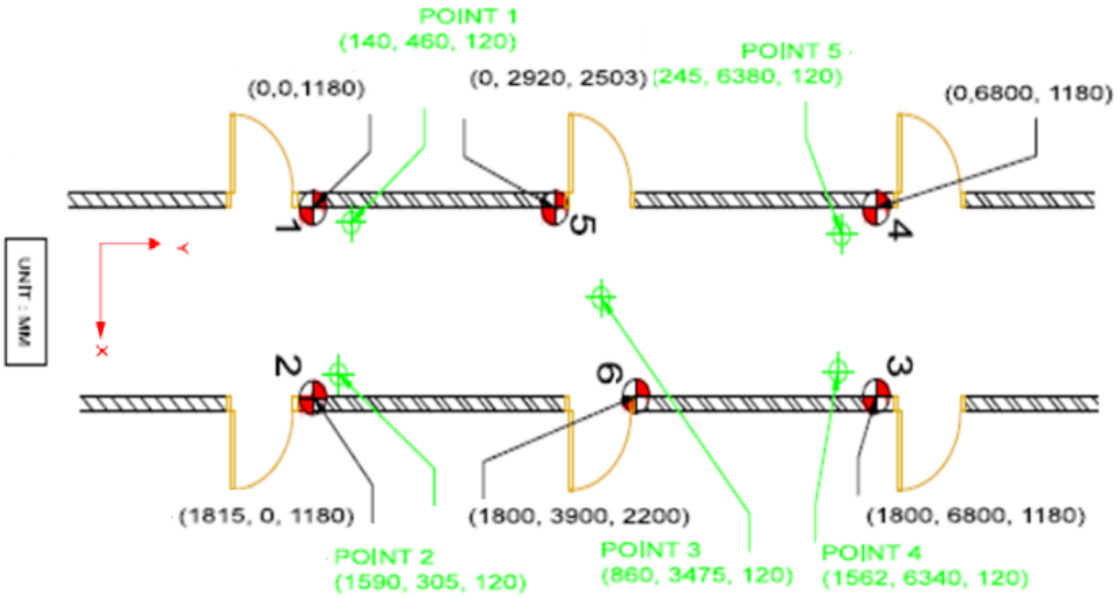


Figure 4.35: Floor Plan of the narrow corridor for UWB positioning test.

The results are shown in the following Figures and summarized in Tables 4.9 and 4.10.

Table 4.9: Statistic analysis on 2D positioning error.

STATISTIC	**UWB-ONLY**	**LMS**
Min [mm]	12.04	1.71
Max [mm]	494.69	340.70
Mean [mm]	145.58	142.52
Median [mm]	143.99	146.62
St.D [mm]	69.78	78.58

Table 4.10: Statistic analysis on 3D positioning error.

STATISTIC	UWB-ONLY	LMS
Min [mm]	23.87	25.79
Max [mm]	499.62	426.82
Mean [mm]	191.82	183.37
Median [mm]	191.61	182.52
St.D [mm]	85.56	91.22

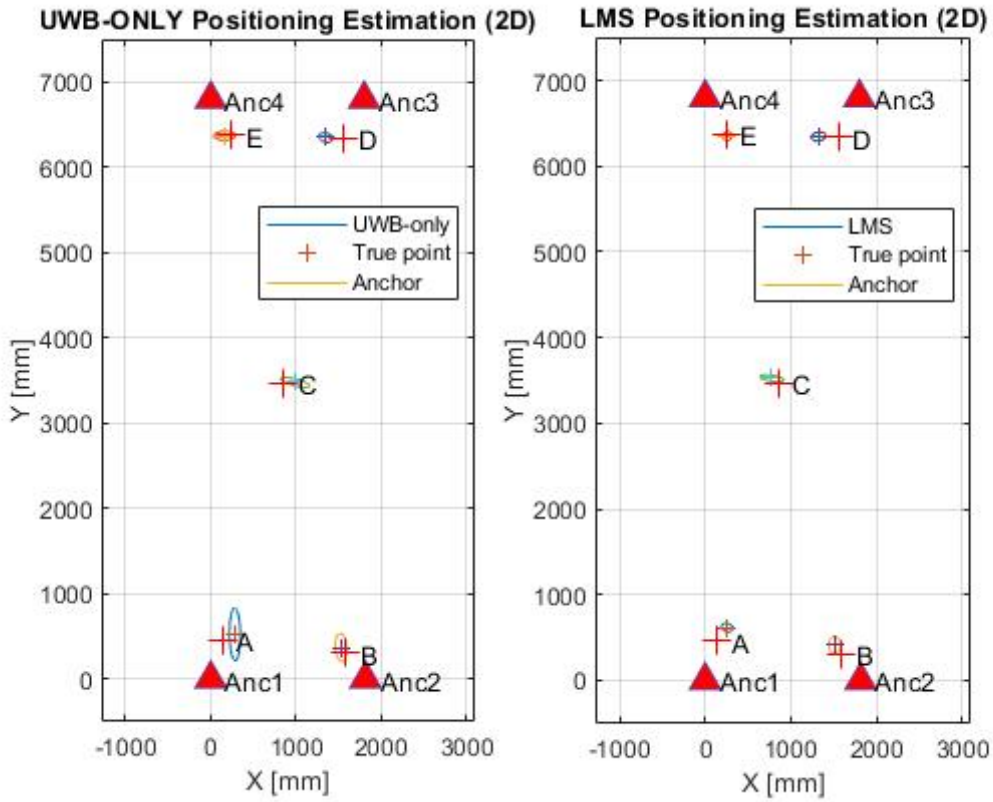


Figure 4.36: Floorplan representation of 2D positioning.

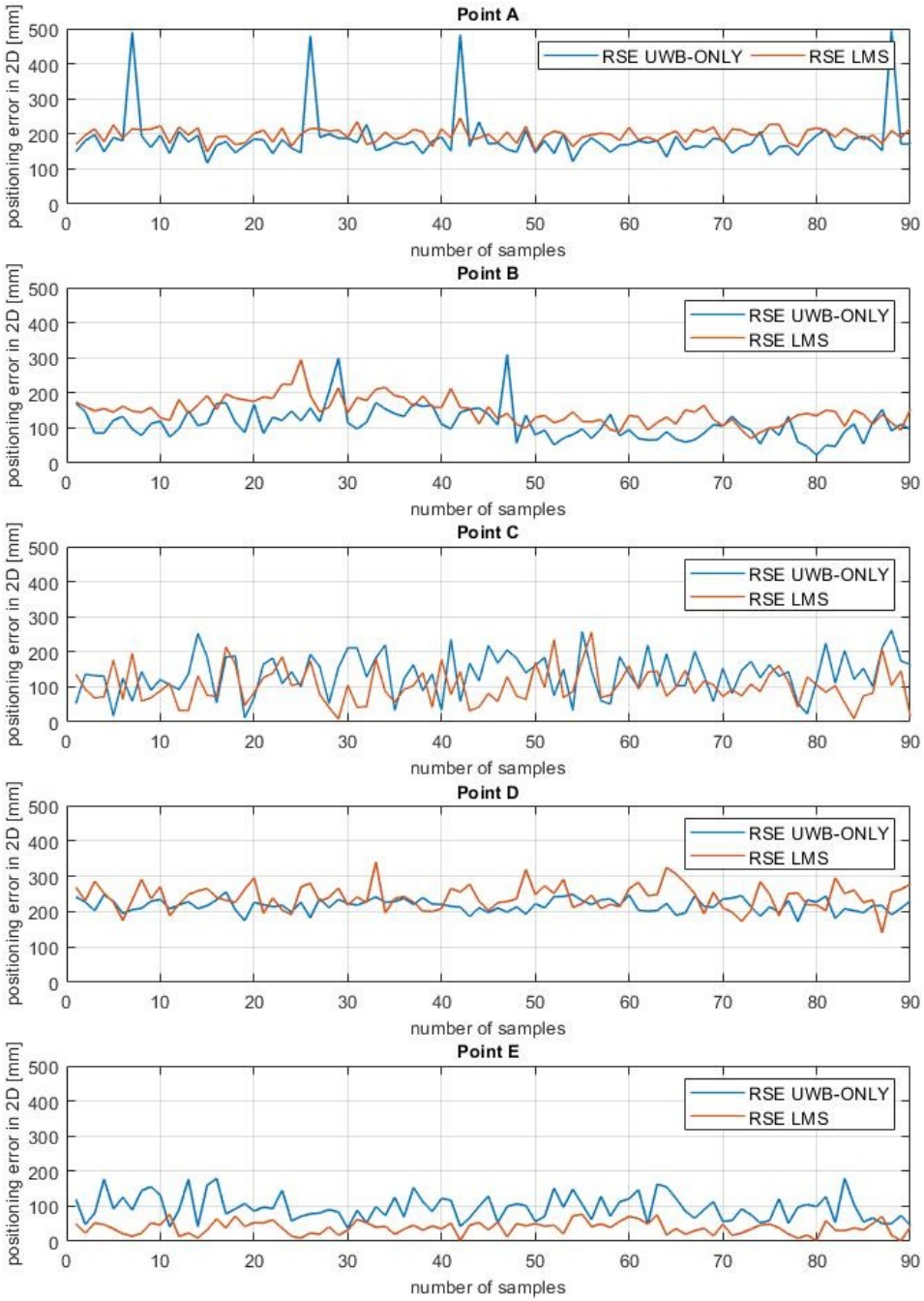


Figure 4.37: Positioning error comparison in 2D.

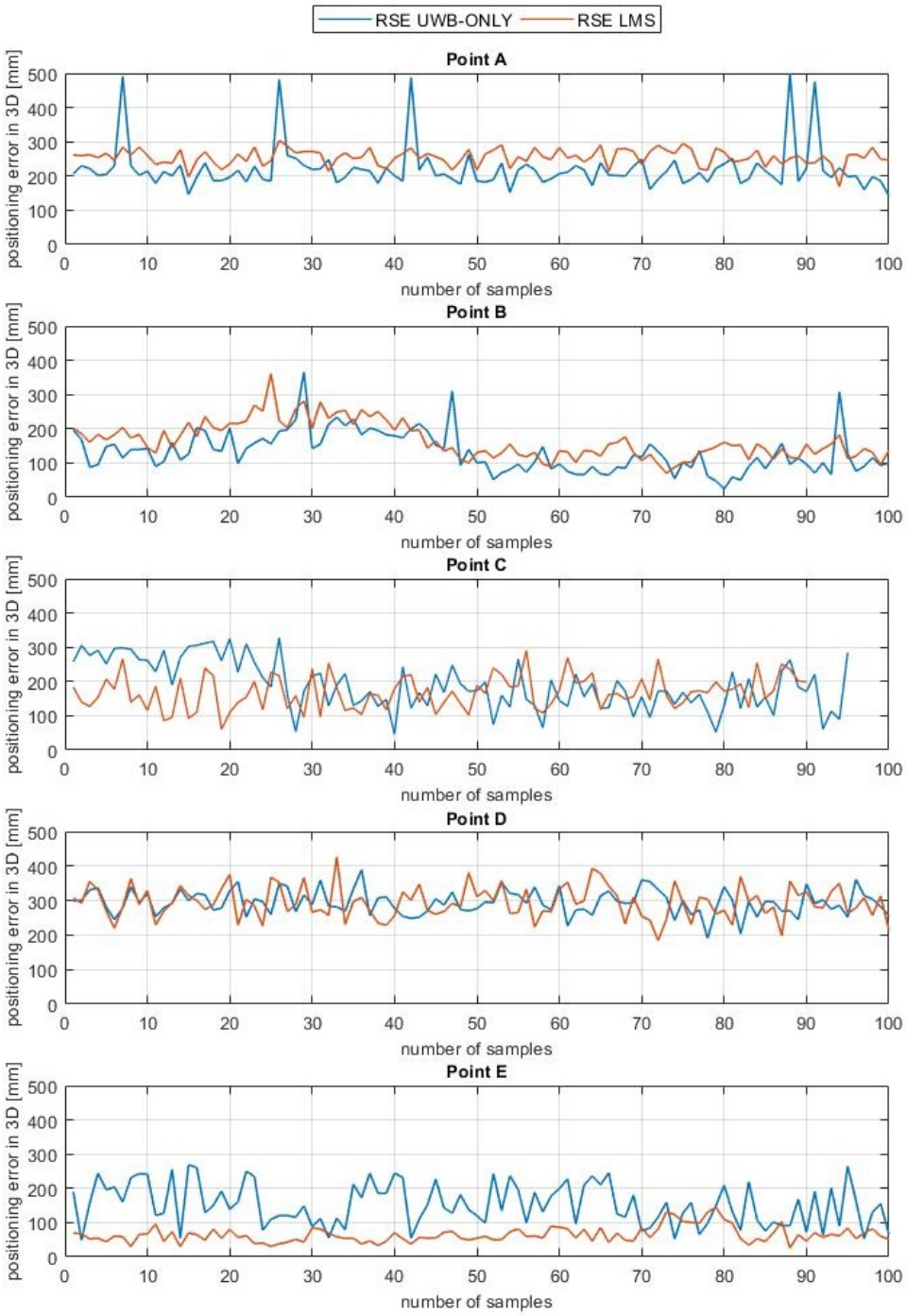


Figure 4.38: Corridor positioning error comparison in 3D.

Again the CDF of the positioning error is reported:

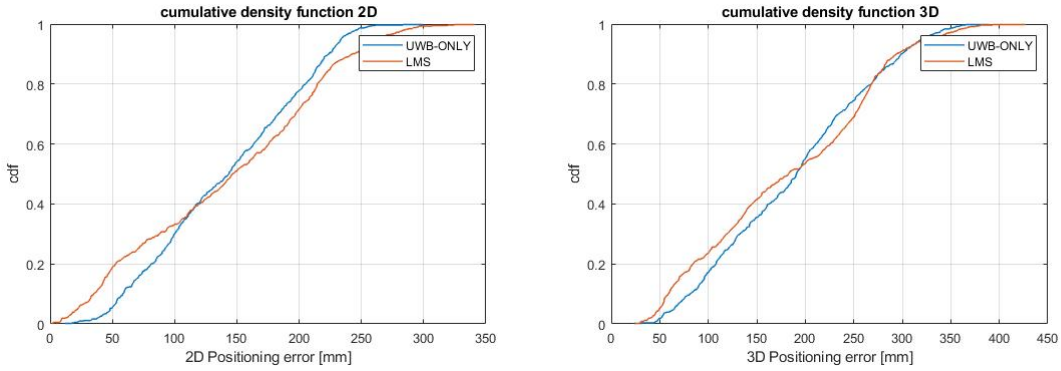


Figure 4.39: CDF of the positioning error in 2D and 3D

4.1.2.5 Outdoor positioning results

The final positioning test was performed in outdoor space with the TREK 1000 evaluation kit. In an area of 100 square meters on a rooftop of Politecnico di Torino, 4 anchors has been located above some tripods forming a network inside which a grid with 2 meters step has been materialized. On the intersection of the grid, 32 points has been accurately measured with a Total Station. The TREK 1000 UWB sensor has been located in static condition for about 2 minutes on each of these points and ranging measurements from the four anchors has been acquired. On these measurements a non linear least square minimization procedure has been used to perform accurate positioning from the range measurements and then to compare the results with the ground truth positions. Figure 4.40 shows the obtained results for this configuration, enhancing the Z component estimation error. Figure 4.41 shows the 2D positioning estimation with respect to the real values measured with the total station.

What is interesting to investigate in this test is the geometric variation of the estimation in function of the position of the UWB receiver. In order to do this, for each of the 32 points of the test, the GDOP value has been computed and then plotted to generate a 2D geometric map (Figure 4.42).

4.1.3 Conclusions

To validate the accuracy and precision of UWB ranging capability and positioning estimation, several tests have been performed with two commercial UWB sensors (Pozyx and DecaWave).

In particular, two sensors has been used to acquire several ranging samples at different distances for both the Pozyx and the TREK 1000 systems. From the analysis of the ranging measurements acquired, it possible to affirm that under 100 meters distance, both systems have a ranging accuracy of about 10 cm with a standard deviation of few centimetres. The distribution of the measurements is

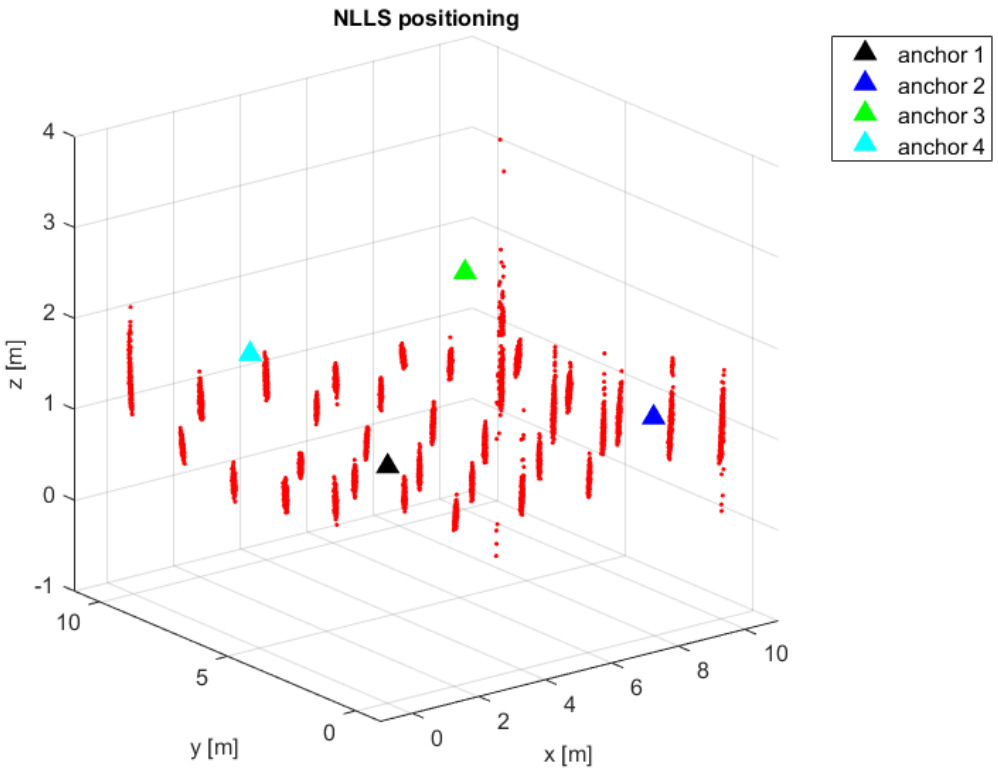


Figure 4.40: NLLS positioning estimation for TREK 1000 outdoor test.

mainly gaussian for LOS condition, while moves to a gaussian-mixture in NLOS conditions.

The Pozyx UWB systems integrates an IMU platform which allows to perform navigation algorithms taking into account the motion of the board. The sensitivity of the IMU has been evaluated comparing the attitude estimates with an higher grade IMU platform and a Total Station as reference solution. The results have shown that the system estimate angles with the typical errors of a consumer grade platform and therefore usable to increment the positioning estimation for example in a Kalman Filter implementation.

The Pozyx system has been used also in indoor scenarios to estimate the positioning capabilities changing the anchor configuration, the type of environments, the surrounding conditions. Several tests has been performed and both inner and proposed algorithms has been used and compared. In general, the positioning performances respect the ones present by the manufactures (accuracy of 10 cm $\pm$ 5 cm). The challenging encountered are mainly related to the presence of strong noise from metallic surfaces or strong reflections from glass walls.

Finally the TREK 1000 system has been evaluated in outdoor condition with a

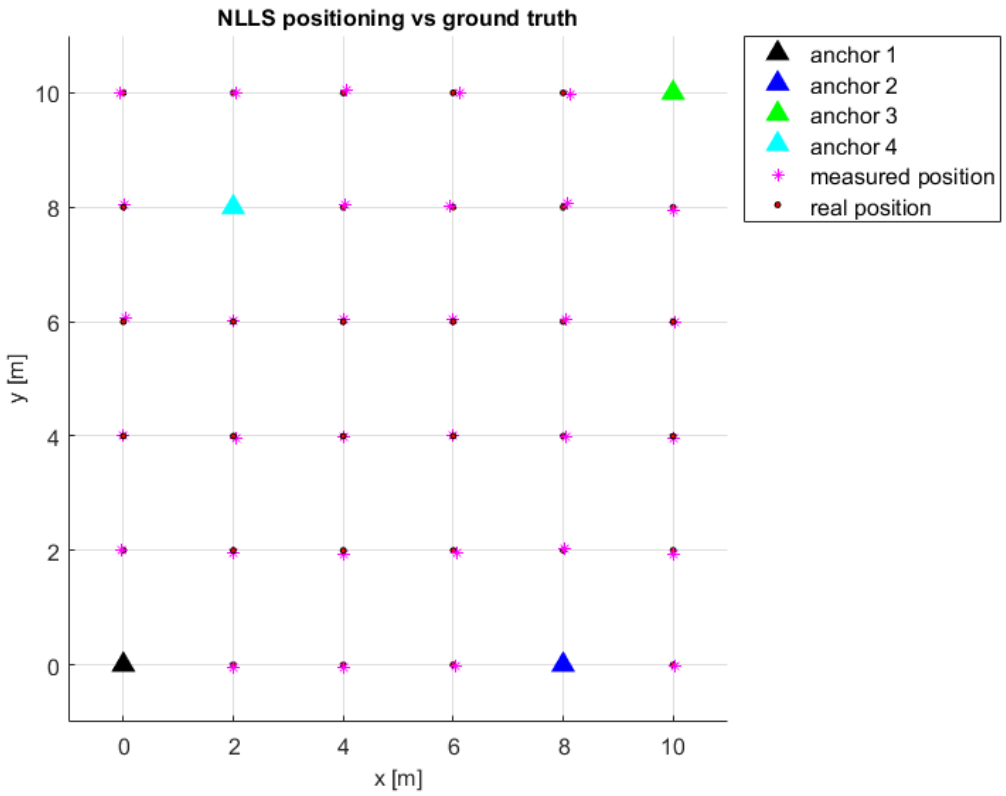


Figure 4.41: NLSS 2D positioning estimation compared with true values from Total Station survey.

proposed positioning algorithm and the GDOP of the geometry has been computed.

Thanks to all these tests is possible to affirm that UWB technology is a suitable solution for positioning and navigation in indoor spaces, like rooms and office floors. This not only thanks to the characteristics of the UWB signal but also thanks to the possibility of scaling the system. A main advantages is the possibility to georeferencing the Anchor infrastructure in order to estimate the receiver position in a geographic reference frame as the GNSS.

Further research on UWB performances should evaluate the multifloor positioning capabilities and the characterization of the NLOS in the signal received by the sensor.

4.2 IRBL in infrastructure scale environment

In the previous section it has been demonstrated how is possible to solve the positioning and navigation problem in indoor environments with a relative low

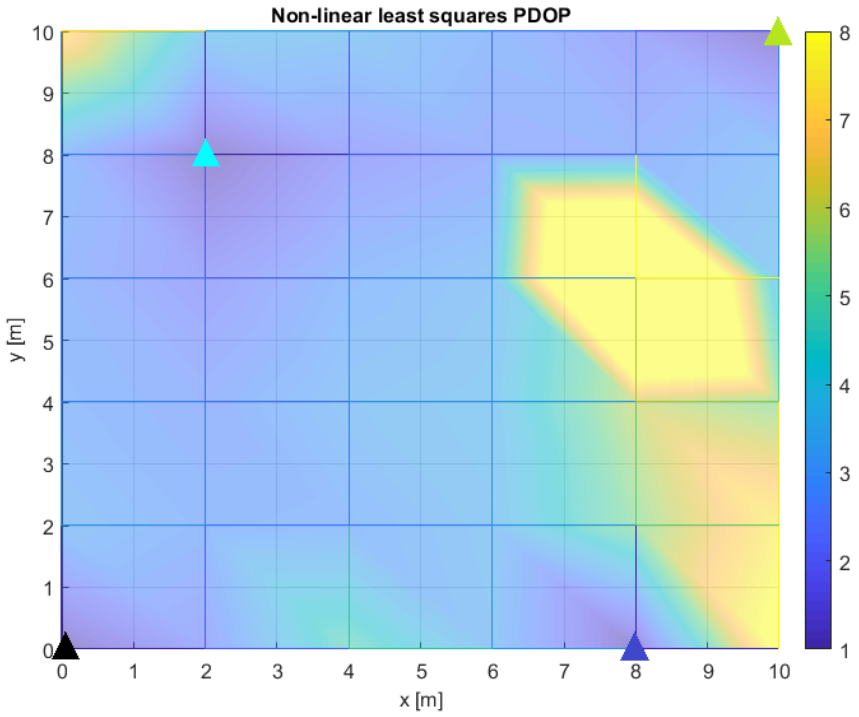


Figure 4.42: GDOP 2D map.

cost and with high accuracy. The seamless positioning requires to estimate the pose of the user in every situation, also when an infrastructure based system like the UWB system cannot be used. Moreover, UWB can't be used in very large environment where the dislocation of the Anchor could requiring numerous sensors and increasing the cost. Image based positioning is a suitable solution for these situations. In particular the proposed solution has been tested in the Infrastructure Scale Environment as defined before. The characteristics of these environments are the large scale, the presence of numerous people and the indoor condition. This section presents an Image based positioning solution developed by the Geomatics Lab's researchers of Politecnico di Torino and validated by the authors in real case studies. Some concept about the procedure has been already presented in Chapter 3 "Image Recognition Based Location (IRBL) Positioning" section.

The IRBL positioning system presented in this work has been applied during two research projects called "A solution of Image Recognition Based Positioning using database of solid images" and "App 4 IRBP Application for Image Recognition Base Positioning". These research projects have been made in collaboration with the Positioning/Navigation Technology Research Section of Electronics and Telecommunications Research Institute (ETRI) Daejeon – Republic of South Korea . The aim of the collaborative project was to implement, test and validate the

Image Recognition Based positioning methodology in Korean indoor test-beds.

The proposed approach of IRBL positioning is based on the use of a single mobile camera as acquiring sensor, under this constraint, in order to estimate the camera parameters (position and orientation), a prior knowledge of 3D environment must be available, in the form of a database of images with associated spatial information. A Terrestrial Light Detection and Ranging (LiDAR) Survey (TLS) with an integrated camera can be used to acquire the 3D model of the environment which is used to generate the images database. The database of synthetic images contains all the information useful to estimate the position and orientation of any camera sensor acquiring the same scene. This thanks to the geometric relations already presented in Section 3.2.2. One of the main challenges of the procedure is to retrieve the correct correspondent image from a database containing millions of images. In this context MPEG algorithms for visual search play an important role in defining light and interoperable solution for processing and comparing the query and database images. The retrieval of the correspondent DB image permit to apply the IRBL algorithm and finally to extract the position and attitudeof the mobile camera image with a photogrammetric approach (Figure 4.43). In contrast with other IPS, the IRBL provides also the camera orientation, a fundamental information for numerous applications like Augmented Reality.

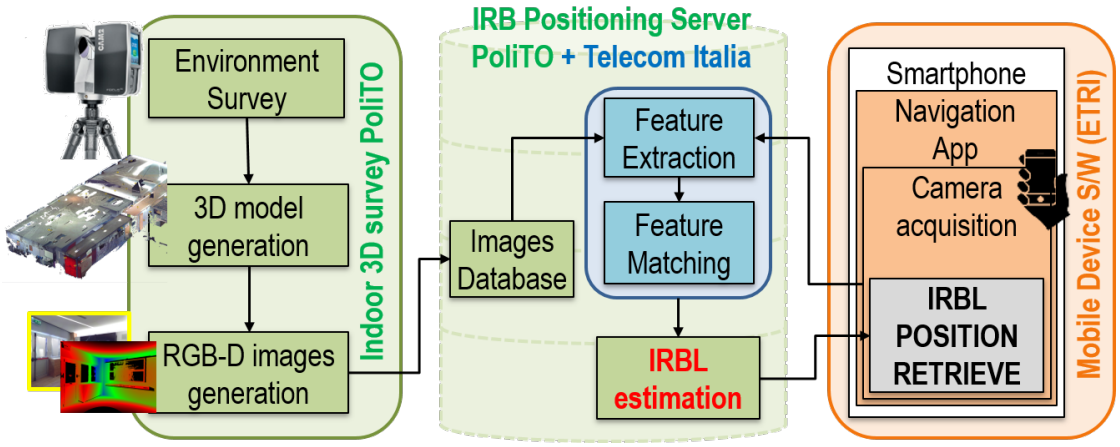


Figure 4.43: The three main core of the IRBL procedure. The DB generation, the server based communication network and the Location algorithm.

The location methodology proposed in this work consists of the following three parts (Figure 4.44):

1. **A methodology for the generation of a RGB-D image DB for object area description:** accurate and exhaustive survey with different techniques in order to generate a 3-dimensional model of the area where the positioning

service is offered. The model is used as input data in an algorithm for the generation of a synthetic images database with related 3D information;

2. **A visual search technology:** in this study, a mobile phone takes a query picture used for locating the camera and so the user; a reference image, that is the most similar to the query one, is extracted from the database thanks to the Compact Descriptors Visual Search (CDVS), patented by TELECOM;

3. **A proposed algorithm for image recognition based location:** IRBL algorithm is based on a sequence of feature matching and robust outlier rejection able to extract a set of 2D feature, homologous points between reference and query images. This 2D features can be transformed in 3D using Solid Image data for a final photogrammetric space rejection.

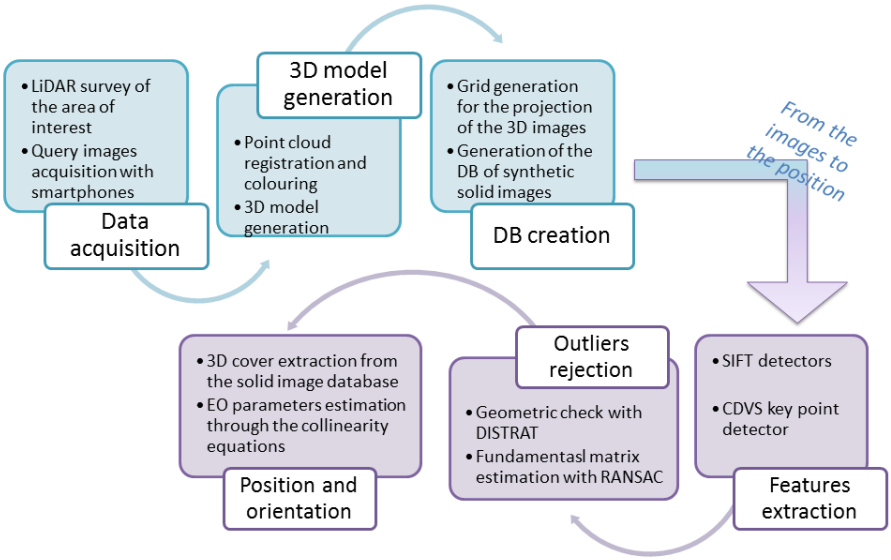


Figure 4.44: The IRBL approach workflow.

The proposed procedure was implemented in a LBS in the form of a smartphone application. The developed application allows the user to acquire an image with his smartphone and receive in real time his position within the selected area. The application was developed by ETRI developers for the Korean territory and is represented in Figure 4.45.

Figure 4.45: Smartphone application based on IRBL procedure.

4.2.1 The 3D model and the synthetic images database

The database of RGB-D images containing image feature with related 3D information is created processing a coloured 3D model of the environment. This model could be extracted from an existing 3D model, generated by a terrestrial or aerial survey, or obtained through a mobile mapping system. In this work the 3D model has been generated with a Terrestrial Laser Scanning (TLS) system, that also allows the acquisition of images using an integrated camera. The acquired point cloud are coloured using the camera associated to the LiDAR instrument. A plurality of different scans are acquired and co-registered. As result of the process, a geo-referenced RGB point cloud of the environment is made, on which is possible to directly read 3D coordinates/color of object points. Once the 3D model is generated, the RGB-D images can be automatically realised by means of the software ScanToRGBDImage developed by the Geomatics group of the Politecnico di Torino.

A RGB-D image is a classical RGB digital image with known internal and external orientation parameters, where a distance between the projection centre and the acquired objects are recorded for each pixel. Therefore, distance values are stored in an additional matrix with the same pixel size, number of columns, and number of rows as the RGB matrix. Additional radiometric information such as NIR, MIR, TIR, multispectral, or hyperspectral bands can be added in other matrix levels. Figure 3.15 in Chapter 3contains a schema of the RGB-D structure. With the generation of a database of RGB-D of an indoor environment, it is possible to

correctly represent the reality.

The software can generate an RGB-D image that needs to contain the following information:

- The external orientation parameters corresponding to the position and orientation of the camera $(X_0, Y_0, Z_0, \omega, \varphi, \kappa)$, which are derived from the position of the point cloud.
- The internal orientation parameters corresponding to focal length, the principal point position of the camera (f, ξ_0, η_0) and distortions (the generated images are synthetic and are considered without distortion).
- The number of pixels in the columns and the rows of RGB-D (n_{row} and n_{col}) and the image pixel size d_{pix}.

As input parameters, the realised program requires the focal length of the images that will be realised, n_{row} and n_{col}, the pixel size of the generated images, and the number of images that need to be extracted according to the vertical (nV) and horizontal (nH) steps (Figure 4.46).

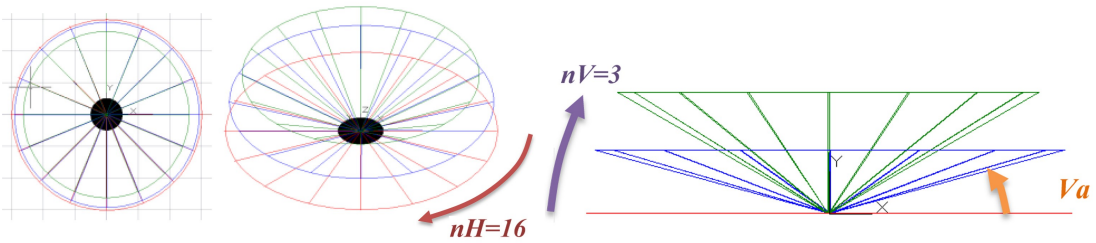


Figure 4.46: An example of definition of RGB-D axis directions for each position.

Once the input parameters are fixed, the process executes the next steps:

1 An empty image (RGB and range matrix levels) is generated using (n_{row}, n_{col});

2 A subset of coloured points (X_i, Y_i, Z_i) with $I = 1 : n$, (n = number of selected points) can be extracted from the original RGB point cloud according to a selection volume that can be defined by a sector of a sphere (Figure 4.47) with:

   – the centre in the location of the generated RGB-D image;
   – the axis direction coincident with the optical axis of the synthetic image;
   – the radius R;
   – the amplitude defined by an angle ($\leq 90°$) that is half of the cone angle measured from the direction axis.

3 For each selected coloured point, a distance di with respect to the location of the generated image is calculated:

$$d_i = \sqrt{(X_i - X_0)^2 + (Y_i - Y_0)^2 + (Z_i - Z_0)^2)} \tag{4.1}$$

4 Each selected RGB point is projected on the synthetic image defining its image coordinates (ξ_i, η_i) by means of the internal and external orientation parameters inside the collinearity equations 3.2.2, where r_{ik} are the coefficients of a 3×3 spatial rotation matrix depending from the camera attitude $(\omega, \varphi, \kappa)$;

5 The image coordinates (ξ_i, η_i) are converted into pixel coordinates (c_i, r_i) using:

$$c_i = \frac{\xi_i}{d_{pix}} + \frac{n_{col}}{2} \tag{4.2}$$

$$r_i = \frac{\eta_i}{d_{pix}} + \frac{n_{row}}{2} \tag{4.3}$$

6 The RGB values of each point are written inside the cells of the image matrix in the position (c_i, r_i);

7 The distance value d_i is written inside the cell of the range image matrix in the position (c_i, r_i).

At the end of the procedure, pixels that are still void are filled by means of an interpolation algorithm based on the nearest filled pixels.

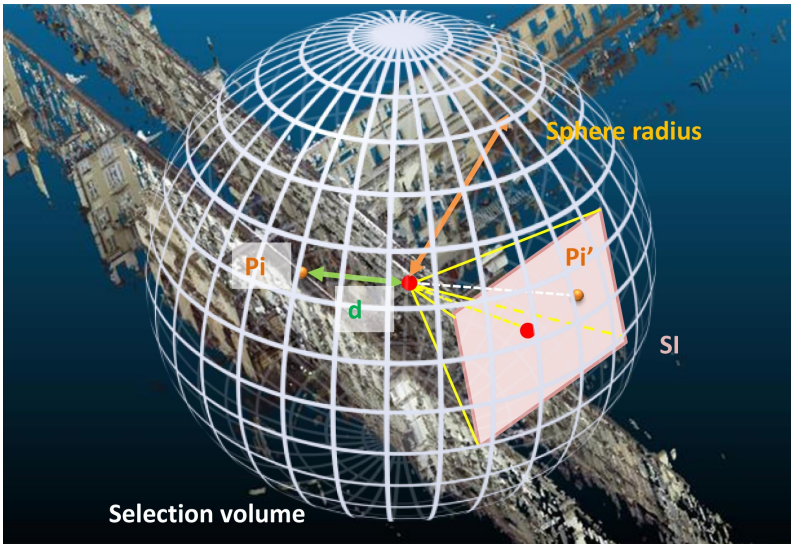


Figure 4.47: The selection sphere for RGB-D image generation.

After the process, ScanToRGBDImage generates a set of synthetic images with the information regarding the position, attitude and ranges i.e. the RGB-D images database.

4.2.2 The Compact Descriptor Visual Search

The goal of the retrieval procedure is to select a reference image out of the images DB with the highest level of similarity with the image acquired by the terminal camera, target of positioning procedure. For the retrieval procedure we adopt the solution defined by MPEG7 CDVS (Compact Descriptors for Visual Search) [73] with minor optimization. To select out of a DB the most similar image the following operations have been defined by MPEG7:

1. local descriptors in query and database images are extracted and compressed;

2. the images are preliminarily ranked based on global descriptor [74] similarity scores between the query image. Global descriptors provide a statistical representation of a set of the most significant local descriptors extracted from the two images. As a result of the global descriptor preliminary screening, several potentially similar images are then selected out of the DB;

3. for the selected best ranked images by the global descriptor similarity test, the pairwise matching procedure between the extracted key points in a couple of images is executed, trying to match similar key points present in both images. For each feature descriptor of the query image, one and only one similar feature descriptor is searched in each single image part of the DB;

4. the matched key points are validated through a geometry check based on the concept that the statistical properties of the log distance ratio for pairs of incorrect matches are distinctly different from the properties of that for correct matches.

   Based on a statistical model, a set of good matches can be ranked using a similarity score given by:

5. the number of correct pairwise key points from the DISTance RATio coherence (DISTRAT) check; and

6. the reliability of each selected match is given by the distance ratio between the first and the second closest descriptors detected in the reference image.

Due to the potential large number of images in the DB, to speed up the retrieval process, CDVS uses compressed descriptors [75]. For this reason, only a limited number of key points are used in the image search procedure. Moreover, the CDVS gives more priority to the points located in the centre of the image. It is evident

that, in some common view, the centre of the picture represents the infinite point of the prospective view so the selected points could be far away from the camera, causing a loss of accuracy in the next step of location. To enhance the accuracy level of the location procedure, the criteria for ranking and selecting the key points should be modified. There is a need for homogeneous distribution of key points in the overall picture, not giving priority to those concentrated in the image centre.

4.2.3 The IRBL algorithm

Once the retrieval of the reference image is completed, it is possible to extract the 3D information of the selected features from the image to estimate the external parameters (position and attitude) of the query image. In details, the 3D information of the reference image is stored inside the DB of RGB-D images where, for each pixel, the distance (range) of the obstacle depicted in the image is reported, together with internal and external orientation parameters (IO/EO). After the extraction of the reference image, the key points and related features are extracted from the query and reference image using a state-of-the-art solution [33] that allows a preliminary association between key points of the two images. After that, a high percentage of outliers rejection is executed according to a new proposed two-step approach. At first, good matches are selected using the DISTRAT algorithm [76], [77] using a geometric check based on the distances ratios between pairs of points in the two analysed images. Then, a RANdom SAmple Consensus (RANSAC) check is executed over a quality improved set of matches. The proposed outlier rejection approach, when applied to real working conditions, reduces the processing time by a factor of 10, with respect to the use of a standard RANSAC approach [34]. Finally, camera parameters are estimated based on 3D information available on the reference image for the selected set of key-point pairs according to the collinearity equations [78]. To analyse the detail of the processing, the next list specifies each step of the IRBL algorithm:

1. extraction of features from query and reference images using scale-invariant feature transform (SIFT) detector;

2. key-point matching procedure where the only query image key points that have one and only one similar descriptor among key points in the reference image are selected, according to SIFT or ORB algorithms;

3. a geometric check (DISTRAT) is used for a coarse preliminary rejection of matched outliers, the use of DISTRAT is required to speed up outliers rejection procedure. On recent implementation of IRBL in C++ version, the DISTRAT is not computed;

4. given the set of common features selected out of the DISTRAT geometric check, the fundamental matrix between the query image and reference image

is estimated with a RANSAC procedure, allowing exclusion of the remaining outliers from the DISTRAT check. The RANSAC is a robust iterative method to estimate the parameters of a mathematical model from a set of observed data that contains outliers, as in the DISTRAT output, a small percentage of outliers are present in the selected set of common features. The RANSAC is a non-deterministic algorithm in the sense that it produces a reasonable result only with a certain probability, with this probability increasing as more iterations are permitted. The preliminary use of DISTRAT reduces the percentage of outliers from 70% to just a few per cent, this allows us to dramatically reduce the RANSAC execution time, by approximately 100 times (at this stage, the focal length is assumed to be similar in both images from the retrieval step, and the camera distortion model is not considered);

5. the common features between the query and reference image are transformed into 3D information using the Solid Image derived from the three-dimensional 3D model of the scene;

6. to improve the initial external and internal orientation parameters of the query image, a direct linear transformation (DLT) could be estimated using the 3D features extracted in the previous step [79]. In recent optimized solution of IRBL, the approximate solution can chose according to the external orientation parameters of extracted Solid Image of DB;

7. rejection of outliers not detected by Steps 3 and 4 are processed by a data snooping process [80]. For the given 11 DLT estimations, the post-fit residuals are calculated in terms of the distance between the projection of the solid point on the query image pair and the matched key-point coordinates. If the largest residual exceeds a threshold, the worst point is discarded and the DLT parameters are estimated again. In recent optimized solution of IRBL, this part is not realized;

8. using the collinearity equations in a least square estimation and the Levenberg-Marquardt method the EO parameters are refined.

The reliability of the final estimated location can be validated using the variance covariance matrix of least square adjustment [81] and checked against the post-fit residuals. This algorithm has been implemented in the MATLAB environment.

4.2.4 3D model generation

The research project between Politecnico di Torino and Electronic and Telecommunication Research Institute is based on the validation of the proposed procedure on two different test sites that have been chosen to have two different indoor scenarios with some specific issues. The first environment, the Bangbae metro station of

Seoul (Republic of Korea), is an important public infrastructure of interest, where an LBS can better express its usefulness. It presents various indoor spaces with different furniture but also a very repetitive railway floor. It is also very populated, which is an important issue in a IRBL system. The second test site is the research department of the ETRI building in Daejeon (Republic of Korea) where, according to the function (research office), the internal areas are repetitive. Each floor has the same aisle with the same colour and the same furniture. The reason for the different scenario is based on the evaluation of the procedure of indoor localisation in noisy areas (very popular with a lot of people), and in similar areas where, from a first view, is difficult to find differences between the different floors (Figure 4.48).

(a) (b)

Figure 4.48: (a) An indoor Bangbae station view; and (b) a typical aisle in the ETRI building.

From the operative point of view, the first step of the work was the realisation of a complete and accurate reference model of the two test areas using a traditional topographic survey and LiDAR acquisition techniques [82]. The topographic survey allows to co-registered the LiDAR scans and to georeference them in a global reference system with cm-level of accuracy. Unfortunately, this procedure has been used only in Bangbae metro-station case study. The ETRI building survey was performed directly with LiDAR acquisition in relative reference system. In this case, the analysis on the following test are still valid as the positioning estimation of the IRBL algorithm will be compared with a ground truth in local coordinate frame. To guarantee continuity of the data in all the environments, several images for a typical photogrammetric approach based on structure from motion (SfM) algorithms were acquired with the idea of combining the data in case of loss of information. As the LiDAR acquisition was suitable for the entire representation of the two environments, the photogrammetric elaboration was not used for the generation of the RGB-D database.

1.2.4.1 Bangbae metro-station

The Bangbae station is a two level metro sation on the Seoul Subway Line 2. It is an underground structure located in Bangbae-dong, Seocho-gu, Seoul. For his 3D model geo-referencing, a mixed GNSS and total station (TS) survey strategy was employed. The network was realised on three main levels of the subway station. The GNSS measurements were acquired outdoor on two vertices connected to Levels -1 and -2 with traditional TS measurements. For the GNSS survey, two Geomax Zenith 35 receiver were employed, and for the TS network, a Leica TS06 was used. In post-processing, the network has been adjusted with Leica Geo-office and Microsurvey Starnet software using the GNSS permanent station of Suwon (a station of the International GNSSs Service network) as reference point 4.49. According to the achieved accuracy on each vertex (less than 1 cm), the next step was the survey of the markers positioned on the station area. This operation was performed with the TS using traditional side-shot measurements. The markers, in this case, black and white checkerboards, are commonly used for the registration of scans and for geo-referencing the final model.

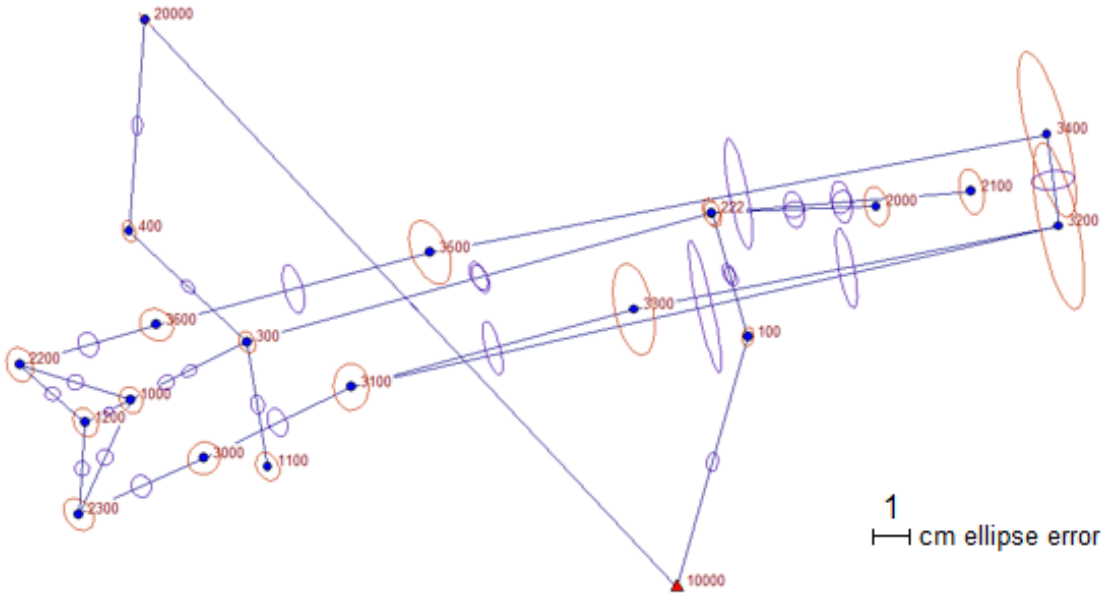


Figure 4.49: The topographic network on floor -1 with errors ellipses after post-processing.

Then two Faro Focus3d X130 Laser scanners were employed. The instrument is a phase shift laser that allows to acquire 3D point clouds with an accuracy of ±2 mm in the range of 0.30–130 m. During the point cloud acquisition, due to the included digital camera, it is possible to acquire the images of the scanned area as well. In the test field, the acquisition was performed with a resolution of 1/5 (a

point each 9 mm at 10 m) and a quality of 4× (points measured four times). For the complete LiDAR survey of the Bangbae subway station, 114 scans were acquired (55 at Level -1 and 59 at level -2). According to the aforementioned setting of the scanner, each scan contains approximately 26 million points, and about three billion points were measured. The LiDAR data were processed using Scene software by Faro.The workflow consist in the following main steps: point cloud colouring, scan registration, and scan geo-referencing. Naturally, using the markers, it is possible to evaluate the accuracy of the geo-referencing according to the residual on the measured point. The mean RMS on the measured markers (85 were employed) was 1.56 cm. Figure 4.50 shows the scan positions while Figure 4.51 shows two views of the complete point cloud (114 merged point clouds).

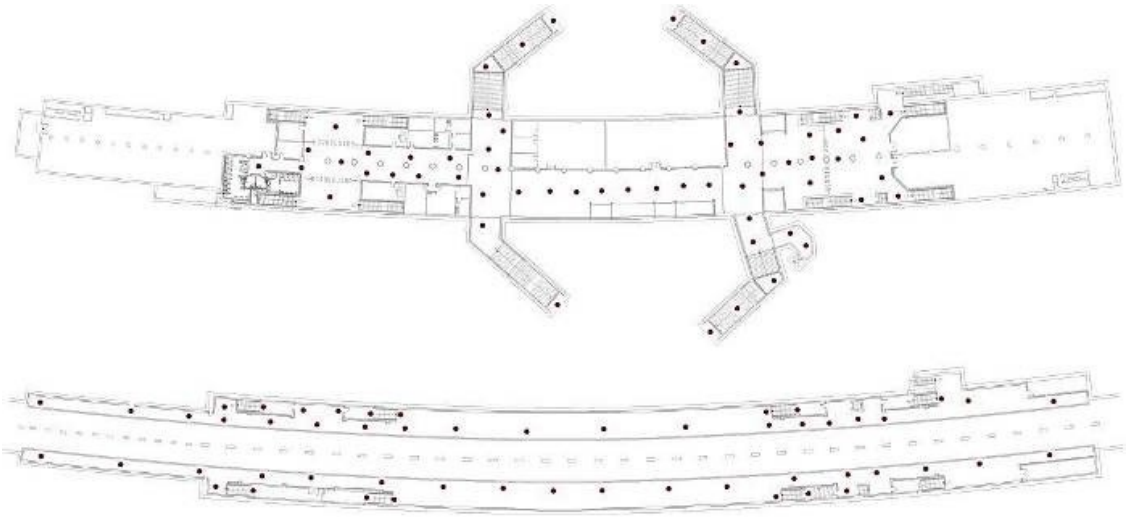
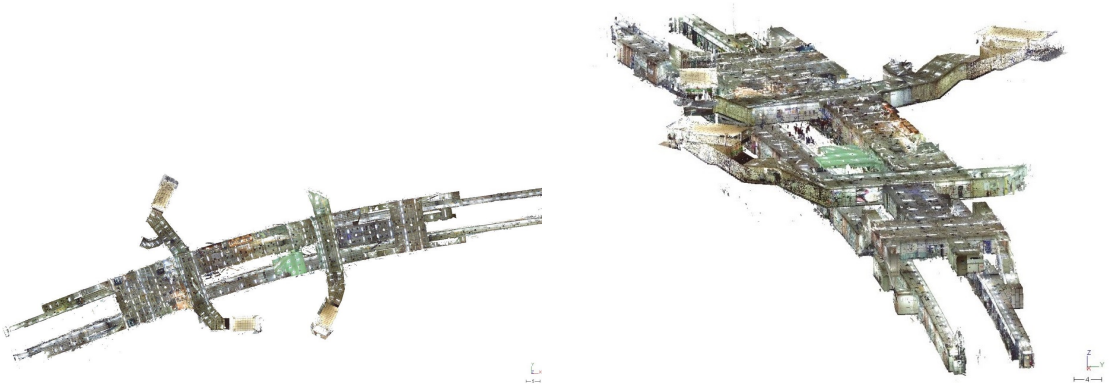


Figure 4.50: Floor plan of Bangbae metro-station and scan position.



(a) (b)

Figure 4.51: An indoor Bangbae station view of the Laser scanner point clouds.

1.2.4.2 ETRI building

The Electronics and Telecommunications Research Institute is a Korean research institution in Daedeok Science Town in Daejeon, Republic of Korea. The headquarters is an office building of 5 floors above the gorund level and one floor underground. The ETRI building was only surveyed by the LiDAR in a local reference system. All the acquisitions were realised without the usually required topographic network and without the markers for the registration of the clouds. As a consequence, the final point cloud is not located in a known cartographic reference system.

As for the Bangbae station, the LiDAR acquisitions have been performed using the aforementioned Faro Focus 3D X130 that was used at a quite higher resolution: 1/4 (a point each 5 mm at 10 m) with the same quality (4×) of the Bangbae settings. The complete building (seven floors) was completely scanned with 111 scans that, according to the setting of the scanner, delivered each scan with 40 million points approximately. Approximately 4.5 billion points were measured (Figure 4.52).

In the case of the ETRI building, the data were processed using Scene software by Faro, but the scan registration was realised using the cloud-to-cloud approach [83]. This approach, based on the well-known iterative closest point (ICP) algorithm [84], has been implemented starting from Version 5.5 of the Scene software and, nowadays is working very well in the pipeline of the Scene LiDAR data processing. Using this approach, it is first important to define an initial setting of the several scan positions. After the initial position, the algorithm allows us to improve the position of the adjacent scans using the shape of the different clouds. In terms of accuracy, in this case, it is possible to understand only the discrepancy between the adjacent clouds that, in the case of the ETRI building, were for all the registered scans under 1 cm. Naturally, as is reported above, with the cloud-to-cloud approach, the geo-referencing was not allowed since no ground control points (GCPs) were measured on the area. All the point clouds were referenced to a local system that started from an arbitrary position of the first achieved scan in the building. In Figure 4.52, a views of the complete point cloud is shown.

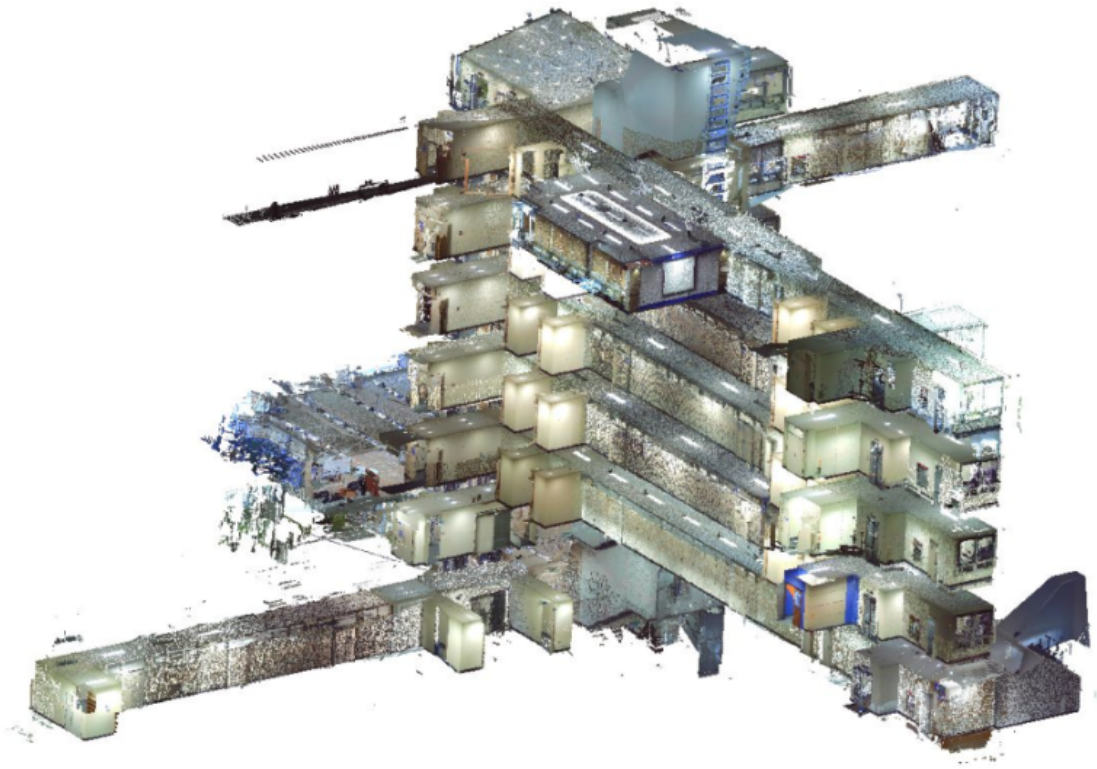


Figure 4.52: LiDAR 3D point cloud of ETRI building.

4.2.5 Synthetic Solid Image database generation

Once obtained the 3D point cloud of the two environment, an ASCII file in .xyz format has been extracted. This data type contains the X, Y, and Z coordinates of each point and the R, G, and B values extracted from the LiDAR internal camera. This file represent the input file for the generation of the RGB-D images. The synthetic RGB-D image can be automatically generated by means of ScanToRGBDImage software tools (developed by the Geomatics research group of the Politecnico di Torino in Intel Visual Fortran) starting from the LiDAR point cloud. The ScanToRGBDImage software generates a set of "synthetic" .JPG images with correspondent range images (Figure 4.53 and 4.54). For each image an orientation file (.ori) is generated with the information about internal and external parameters. This parameters must be set as a-priori information to obtain the best possible image in term of pixel size, field of view and image dimension. For each scan position, 96 images have been generated: 32 horizontal directions for three different inclinations of 0°, 10°, and 20° with respect to the horizontal plane with 2500×1600 pixels, $3\mu m$ pixel size, and a focal length of 4.667 mm. For the Bangbae DB, almost 9700 RGB-D images have been produced in about 36 hours of batch processing time with a desktop computer (i7 5600 U 2.66 GHz 32 Gb RAM), while

for the ETRI building, 10.700 images have been produced in about 40 hours with a computer with the same characteristics.

Figure 4.53: (a) Example of six RGB-D images generated with the software ScanToRGBDImage in RGB visualization; and (b) example of six RGB-D images in a depth map visualisation.

Table 4.11: camera center position and attitude for the previous six images.

Image Name	**Position [m]**			**Attitude [gon]**		
	X	Y	Z	ω	φ	κ
S01	322920.858	4150175.414	45.967	100	0	0
S02	322920.858	4150175.414	45.967	100	-25	0
S03	322920.858	4150175.414	45.967	100	-50	0
S04	322920.858	4150175.414	45.967	100	-75	0
S05	322920.858	4150175.414	45.967	100	-100	0
S06	322920.858	4150175.414	45.967	100	-125	0

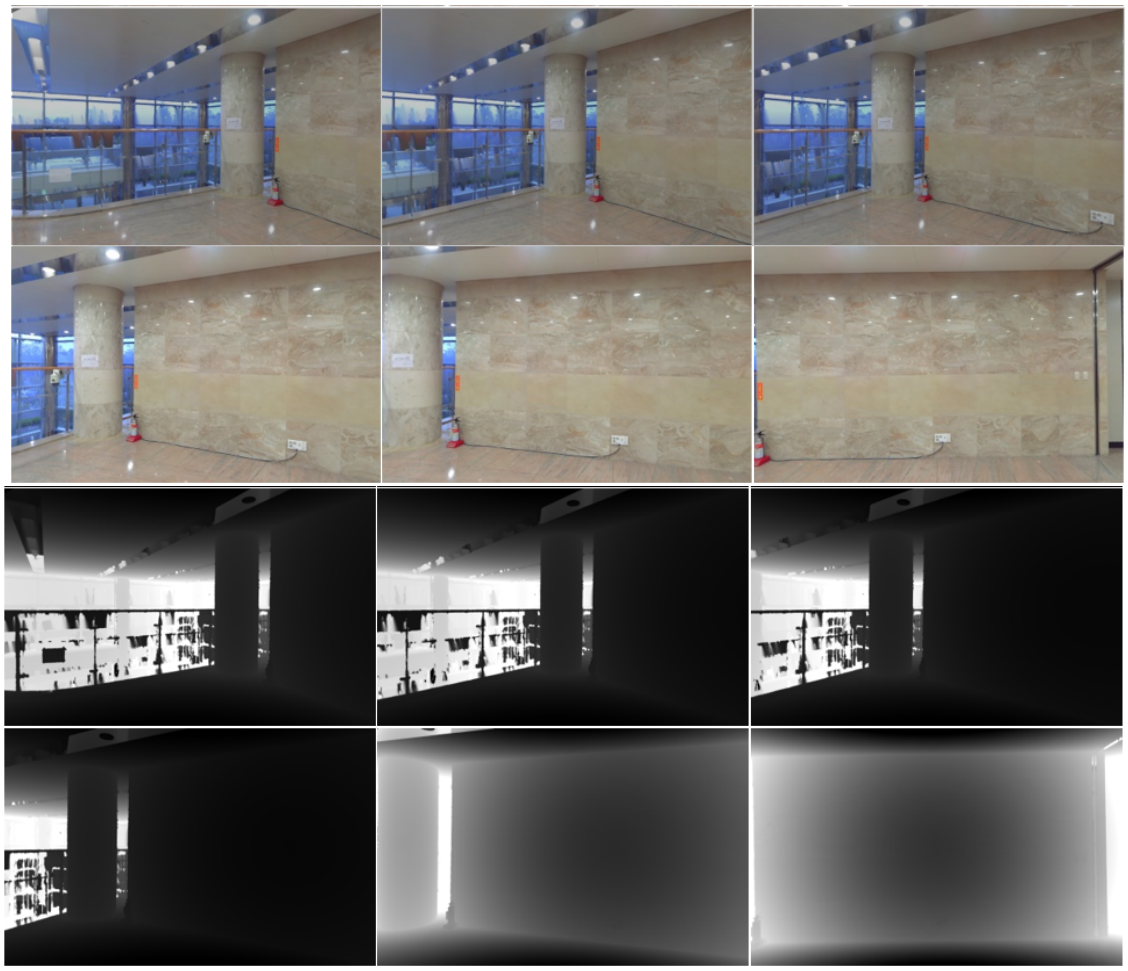


Figure 4.54: (a) Example of six RGB-D images generated with the software ScanToRGBDImage in RGB visualization; and (b) example of six RGB-D images in a depth map visualisation.

Table 4.12: The position and attitude information of the camera center for the previous six images.

Image Name	**Position [m]**			**Attitude [gon]**		
	X	Y	Z	ω	φ	κ
S085_04	-67.20840	-20.76387	97.35983	100	-36	0
S085_05	-67.20840	-20.76387	97.35983	100	-48	0
S085_06	-67.20840	-20.76387	97.35983	100	-60	0
S085_07	-67.20840	-20.76387	97.35983	100	-72	0
S085_08	-67.20840	-20.76387	97.35983	100	-80	0
S085_09	-67.20840	-20.76387	97.35983	100	-92	0

4.2.6 Smartphone image retrieval procedure

The RGB-D images generated with the above automatic procedure has been uploaded in a server machine to form the reference database. Through a client-server internet communication is possible to connect any device to the DB and extract the required information. In particular, the user located in an unknown environment, can acquire an image with the smartphone camera and send it to the correct storage space in the server. The data management and the user interaction is a crucial part of the LBS design as the user must have the access to the correct DB related with his own location. This means that the image acquired with the smartphone camera should be sent to the server together with the information regarding the relative database. Moreover, the IRBL algorithm, running on the server machine, need the internal parameters information of the camera which has shot the query picture. In this case, the data packet should contain also the focal length and the pixel size of the camera. More in details, the code has been designed in order to acquire all this information from the query image name (light approach) which structure is the following:

ExecutionMode_DataBaseName_FocalLength_DimPixel_FileName.jpg

Where:

- *ExecutionMode* is 1 for simplefied procedure, 2 for full IRBL;
- *DataBaseName* is a two characters string corresponding to the name of a Data Base containing the reference images (e.g. BB for Bangbae, EB for ETRI building);
- *FocalLength* is the focal length of the camera used to acquire the image;
- *DimPixel* is the pixel size of the acquired image.

As stated in Section 4.2.2, the visual search technology allows us to retrieve the best reference images form the RGB-D images database and ranked them with a priority score. This procedure was applied on a selected number of query images for both the test sites (20 images for Bangbae and 10 images for ETRI building), and the results of the extraction are shown in Table 4.13 for the Bangbae metro station and in Table 4.14 for the ETRI building. In these tables, the obtained scores of the 1st ranked image selected by the CDVS server are reported. This is the best solutions from the three possible candidates proposed by CDVS. As shown in Table 4.13, the score is always greater than 3, indicating quite good solutions. In most cases, the score is greater than 5, indicating a good solution. The time for the query retrieval process is estimated at about three seconds. In the second test site, 10 check images have been acquired by the smartphone Samsung S7. The results of the reference image extraction using CDVS are greater than 3, indicating

quite good solutions, excluding Image No. 2 (score = 2.54) that was ignored since the resulting IRBL solution was incorrect.

Table 4.13: CDVS image retrieve score. Reference and Query images of Bangbae subway station.

Smartphone image	**Reference image 1**	**Score 1**
query/1.jpg	dataset/b022_i____+0_+0_24_02	8.39
query/2.jpg	dataset/b012_i____+0_+0_09_01	8.91
query/3.jpg	dataset/b002_i____+0_+0_10_01	7.43
query/4.jpg	dataset/b004_i____+0_+0_10_01	17.76
query/5.jpg	dataset/b006_i____+0_+0_07_03	15.39
query/6.jpg	dataset/b013_i____+0_+0_27_01	31.93
query/7.jpg	dataset/b013_i____+0_+0_25_01	9.56
query/8.jpg	dataset/b007_i____+0_+0_16_03	4.52
query/9.jpg	dataset/b011_i____+0_+0_15_02	48.76
query/10.jpg	dataset/b011_i____+0_+0_18_02	20.26
query/11.jpg	dataset/v004_i____+0_+0_15_02	42.23
query/12.jpg	dataset/v020_i____+0_+0_09_01	4.51
query/13.jpg	dataset/v008_i____+0_+0_10_01	3.14
query/14.jpg	dataset/v008_i____+0_+0_25_02	12.05
query/15.jpg	dataset/v006_i____+0_+0_11_01	3.36
query/16.jpg	dataset/v040_i____+0_+0_22_01	6.08
query/17.jpg	dataset/v038_i____+0_+0_04_02	59.79
query/18.jpg	dataset/v038_i____+0_+0_07_01	11.72
query/19.jpg	dataset/v039_i____+0_+0_25_02	3.21
query/20.jpg	dataset/v023_i____+0_+0_26_02	8.83

Table 4.14: CDVS image retrieve score. Reference and Query images of ETRI building.

Query images	**Reference Image 1**	**Score**
query/1__01.jpg	dataset/s020_i____+0_+0_25_01.jpg	7.27
query/1__02.jpg	dataset/s021_i____+0_+0_12_02.jpg	2.54
query/2__03.jpg	dataset/s011_i____+0_+0_12_03.jpg	4.54
query/3__03.jpg	dataset/d011_i____+0_+0_31_01.jpg	6.02
query/3__05.jpg	dataset/d008_i____+0_+0_29_01.jpg	8.51
query/4__01.jpg	dataset/s066_i____+0_+0_18_01.jpg	6.46
query/4__03.jpg	dataset/d008_i____+0_+0_17_01.jpg	6.56
query/5__01.jpg	dataset/d012_i____+0_+0_03_01.jpg	9.6
query/5__04.jpg	dataset/d011_i____+0_+0_19_01.jpg	6.25
query/5__06.jpg	dataset/d012_i____+0_+0_03_01.jpg	8.10

4.2.7 IRBL validation

The same pictures used to evaluate the CDVS image retrieval has been used as ground truth to validate the IRBL algorithm pose and attitude estimation. For this purpose the, camera center position and orientation must be known. As no location information was acquired during the smartphone image acquisition, the external orientation parameters of these images has been estimated with an accurate photogrammetric procedure. It consist in estimate the internal orientation parameter of each smartphone used in the test (five parameters: ξ_0, η_0, c, K_1, and K_2) by a camera calibration algorithm and then evaluate the external orientation parameters (six parameters: $X_0, Y_0, Z_0,\ \omega, \phi, \kappa$) using a single image adjustment (based on pyramid vertex). The camera calibration allows the evaluation of the effects of the radial and tangential distortion of the sensors that are involved in the definition of the camera internal orientation using the collinearity equations. However, as an approximation, it is possible to consider only the effects of the radial distortion, expressed in this case by two parameters K_1, and K_2.

Knowing the object coordinates of some points acquired by the camera, it is possible to obtain the unknown parameters by solving the bundle-adjustment calculation. The object on which the calibration is usually made is a calibration grid where the coordinates of the grid points are known with high precision. This procedure is known as the self-calibration of the camera sensor. The "Camera Calibrator" tool of MATLAB was used. The application requires the use of a specific checkerboard pattern that must not be square. The images of the pattern must be acquired with a fixed zoom and focus. The calibration requires at least three images, but it is suggested to use 10–20 images from different distances and orientations to obtain the best results. A reference system is also defined using the different numbers of squares in the two directions. The calibration algorithm assumes a pinhole camera model, and after processing the applications, displays the results and the accuracies of the process. In this work, the self-calibration was made on the three different smartphones used for the data acquisition and the results are shown in Table 4.15
.

Table 4.15: Internal calibration parameters

Parameters	Samsung Galaxy A5	Samsung Galaxy S5	Samsung Galaxy S7 Edge
Pixel size (μm)	1	1.14	1.4
Focal lengths f_x (pixels)	3706.0	4290.8	3168.7
Focal lengths f_y (pixels)	3722.6	4282.8	3178.9
Princ. Point ξ_0 (pixels)	2070.1	2667.8	1995.3
Princ. Point η_0 (pixels)	1135.4	1477.8	1204.4
Radial distortion K_1	0.1386	0.1148	0.3444
Radial distortion K_2	-0.2587	0.0100	-0.6117
Focal length (mm)	3.714	4.801	4.446
Princ. Point ξ_0 (mm)	0.006	0.014	-0.087
Princ. Point η_0 (mm)	0.026	0.018	-0.179

After the internal calibration, to define the position and attitude of the acquired smartphone images and then use it as "ground-truth", a photogrammetric process was employed. In the case of single-shot acquisition, it is possible to perform single image adjustment (or pyramid vertex) that allows us to evaluate the coordinates of the acquisition point (X_0, Y_0, and Z_0) and the orientation of the camera ($\omega, \phi, and \kappa$). For this task, at least six collinearity equations must be written i.e. three plano-altimetric GCPs are required. The coordinates of the GCPs were extracted directly from the previous LiDAR point clouds using Scene. First, a visible point was selected on the smartphone image. Afterwards, the same point was measured on the point cloud, and the coordinate were extracted. These values (coordinates) were used as GCPs in the employed photogrammetric software (Figure 4.55). Erdas Imagine by Hexagon Geospatial was employed for the process. The final precision for all the analysed images was around 5 cm for the position and around 10 mgon for the angular values. Twenty Query images were used for the check for Bangbae station (10 images for each floor) and 10 images were used for the ETRI building. Tables 4.16 and 4.17 shown the absolute position and orientation of the smartphone images acquired in different location at Bangbae station. For the ETRI building, the results in local reference system are reported in Table4.18.

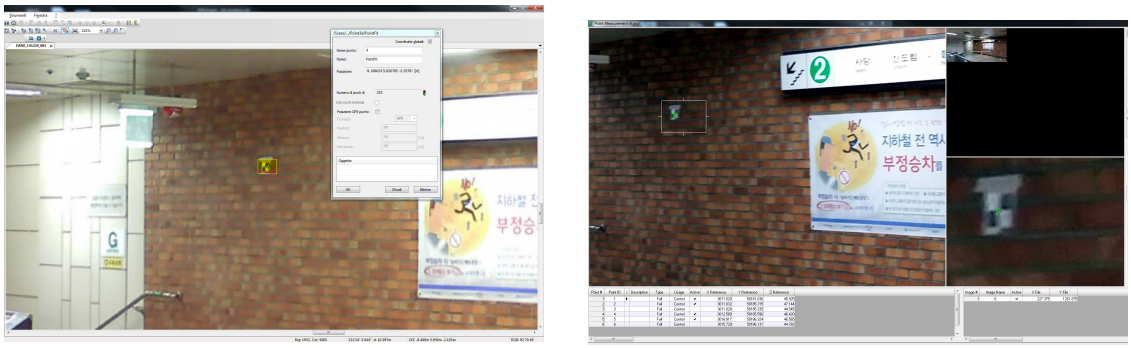


Figure 4.55: (a) GCP coordinate extraction from LiDAR data (a), GCP measurement in ERDAS (b).

Table 4.16: Reference solution with photogrammetric approach. Bangbae subway station main floor.

Image n.	**X [m]**	**Y [m]**	**Z [m]**	ω [gon]	φ [gon]	κ [gon]
1	2988,26	50172,61	45,99	98,327	304,094	200,612
2	2992,38	50174,64	45,94	-75,445	286,702	-174,609
3	3001,95	50175,94	46,02	-15,986	-107,949	-113,858
4	3007,43	50177,38	46,12	-59,216	-110,713	242,225
5	3011,35	50178,21	46,29	137,069	286,305	36,825
6	3021,94	50179,75	46,14	100,344	135,506	-1,131
7	3020,70	50180,34	46,17	219,952	96,320	-120,909
8	3020,95	50175,47	45,76	305,100	-45,201	204,819
9	3020,07	50177,55	46,23	301,999	224,816	201,002
10	3019,666	50177,06	46,101	299,053	172,266	-199,123

Table 4.17: Reference solution with photogrammetric approach. Bangbae subway station train floor.

Image n.	**X [m]**	**Y [m]**	**Z [m]**	ω [gon]	φ [gon]	κ [gon]
11	3034,38	50175,28	41,55	306,159	-53,700	206,084
12	2965,49	50160,10	41,04	-102,177	-116,424	198,612
13	2991,07	50166,68	41,24	-108,802	-111,284	192,534
14	2998,33	50167,47	41,35	122,299	107,500	-219,836
15	3015,30	50171,82	41,43	-85,782	-89,015	214,500
16	2960,70	50171,93	41,02	-244,879	74,385	196,816
17	2984,19	50177,20	41,32	101,869	-55,630	0,233
18	2984,98	50178,40	41,31	76,021	-97,877	-23,253
19	2978,58	50177,46	41,29	99,294	131,964	2,216
20	2900,44	50149,40	40,79	108,389	122,153	391,479

Table 4.18: Reference solution with photogrammetric approach. ETRI building.

n.	**Image**	**X [m]**	**Y [m]**	**Z [m]**	ω [*gon*]	φ [*gon*]	κ [*gon*]
1	1_01.jpg	-72,288	-17,302	110,250	-60,191	105,643	160,590
2	1_02.jpg	-71,776	-17,120	110,283	303,925	-95,921	203,879
3	2_03.jpg	-61,701	-17,518	114,367	310,067	-79,841	211,176
4	3_03.jpg	-61,844	-1,883	117,386	97,494	2,214	-0,013
5	3_05.jpg	-61,769	-10,470	110,196	94,362	6,305	2,316
6	4_01.jpg	-61,560	0,879	110,219	-96,539	31,940	200,618
7	4_03.jpg	-61,961	-9,080	118,367	-99,984	32,258	199,211
8	5_01.jpg	-61,738	4,402	118,608	97,693	-33,397	0,052
9	5_04.jpg	-61,675	4,541	118,626	-97,587	30,747	200,608
10	5_06.jpg	-61,634	2,853	118,621	93,111	-42,916	0,373

Thanks to this procedure, the IRBL positioning and attitude estimation results can be compare with the real camera pose obtained by the more accurate photogrammetric procedure. For the validation, the last version of IRBL code in C++ has been used. In this implementation is possible to select between two different feature extractors (SIFT or ORB).

4.2.7.1 Results and Validation

In this section the results and validation of the IRBL algorithm are reported. The results are referred to the IRBL procedure using SIFT algorithm. The ORB implementation has given similar results and therefore reported only for processing time comparisons. For each test site (Bangabe subway station main floor, Bangbae subway station train floor, ETRI building) the position and attitude of the Query images obtained by the IRBL procedure are presented together with the processing time required. Using the reference solution obtained from the photogrammetric procedure, the discrepances in positioning and orientation along the three axis has been computed and reported. The results shown that:

1. **Bangbae subway station main floor.**
    - The discrepancies $\Delta X, \Delta Y$ are always lower than 1.6 m in absolute value. Excluding the greater values (ΔY for image n. 10), the discrepancies are lower than 1.2 m (4.20);
    - The angular values are estimated with a precision of about 10 gon (mean value).
    - The standard deviations of the discrepancies $\Delta X, \Delta Y$ are of about 70 cm, this is the precision of IRBL positioning in XY using SIFT;
    - The discrepancies ΔZ are always lower than 0.53 m in absolute value;

- The standard deviations of the discrepancies ΔZ are of about 15 cm, this is the quality in Z of the estimated location.

Table 4.19: IRBL estimation for Bangbae subway station main floor.

Time [s]	**Image**	**n. points**	**X [m]**	**Y [m]**	**Z [m]**	ω [*gon*]	φ [*gon*]	κ [*gon*]
4,953	1	156	2988,09	50173,46	45,77	85,620	334,978	209,424
4,755	2	161	2993,43	50174,99	45,63	125,088	315,314	25,894
4,781	3	148	3001,18	50176,02	45,69	185,527	308,452	87,842
4,985	4	142	3007,10	50176,92	45,60	147,638	309,179	49,148
4,703	5	153	3011,21	50178,08	45,79	348,510	283,578	249,019
4,781	6	131	3021,94	50179,70	45,94	301,536	135,809	198,105
4,922	7	142	3021,82	50179,73	46,09	198,937	95,956	300,138
4,890	8	121	3020,84	50175,42	45,96	101,725	373,332	198,649
4,781	9	133	3020,85	50177,42	45,96	100,220	224,143	0,044
4,875	10	147	3020,84	50175,45	45,95	97,580	170,248	0,539

Table 4.20: Discrepances between reference solution and IRBL estimation. Bangbae subway station main floor.

Image n.	ΔX **[m]**	ΔY **[m]**	ΔZ **[m]**	$\Delta\omega$ **[gon]**	$\Delta\varphi$ **[gon]**	$\Delta\kappa$ **[gon]**
1	-0,174	0,855	-0,214	-12,707	30,885	8,813
2	1,048	0,352	-0,310	0,533	28,612	0,503
3	-0,778	0,079	-0,333	1,513	16,401	1,701
4	-0,323	-0,462	-0,522	6,854	19,891	6,922
5	-0,138	-0,132	-0,503	11,442	-2,727	12,195
6	0,000	-0,057	-0,201	1,192	0,303	-0,764
7	1,125	-0,609	-0,071	-21,016	-0,364	21,047
8	0,11	0,05	-0,20	-3,375	18,533	-6,170
9	0,785	-0,129	-0,272	-1,779	-0,673	-0,959
10	1,172	-1,605	-0,150	-1,473	-2,018	-0,338

The results obtained are summarize with their statistical parameter in Table 4.21

2. **Bangbae subway station train floor.**

   - From results reported in Table 4.23 is possible to observe a discrepancy of about 6 meters in X position for image 12. Observing the images is evident that the CDVS procedure has extracted a wrong image which is quite far from the smartphone Query image (Figure 4.56). Although

Table 4.21: Statistical parameters results

	ΔX [m]	ΔY [m]	ΔZ [m]	$\Delta\omega$ [gon]	$\Delta\varphi$ [gon]	$\Delta\kappa$ [gon]
min	-0,778	-1,605	-0,522	-21,016	-2,727	-6,170
max	1,172	0,855	-0,071	11,442	30,885	21,047
mean	0,283	-0,166	-0,277	-1,881	10,884	4,295
st.d.	0,693	0,649	0,145	9,228	13,367	7,998

the retrieving score is good due to the similarity of the environment, the Reference image used in the IRBL procedure is wrong an the result is not correct. In this case, the importance of the DB density in term of images has been highlighted.

Figure 4.56: Wrong images correspondence due to the CDVS retreivial.

- The discrepancies in X and Y are always lower than 1.5 m in absolute value, excluding the gross error of image n. 12 in X coordinate. The shape of train floor (long and narrow) can be underline some critical problem of uncorrect geometry of feature points;
- The standard deviations of the discrepancies in X is about 1 m and in Y is about 50 cm, this is the XY quality of DB (Table 4.24);
- The discrepancies in Z are always lower than 1.20 m in absolute value;
- The standard deviations of the discrepancies in Z are of about 40 cm, this is the Z quality of DB;
- The angular values are estimated with a precision of about 10 gon;
- The estimated averages are not significant for all the parameter, than there are not any systematic error estimation: precision and accuracy are substantially coincident.

Table 4.22: IRBL estimation for Bangbae subway station train floor.

Time [s]	**Image n.**	**n. points**	**X [m]**	**Y [m]**	**Z [m]**	ω [*gon*]	φ [*gon*]	κ [*gon*]
4,953	11	176	3034,11	50175,72	41,61	306,67	-52,10	206,417
4,755	12	22	2971,81	50161,54	41,16	-107,60	-115,10	190,24
4,781	13	17	2989,89	50167,71	41,43	-127,79	-105,94	174,47
4,985	14	62	2999,42	50167,70	41,31	115,98	108,61	-214,394
4,703	15	16	3014,39	50172,31	41,98	-72,85	-85,93	228,197
4,781	16	19	2959,18	50171,22	42,29	-226,35	84,37	177,549
4,922	17	314	2983,93	50177,03	41,33	101,54	-56,28	-0,14
4,890	18	76	2984,25	50178,24	41,26	87,32	-97,25	-12,488
4,781	19	17	2979,17	50177,72	41,23	98,47	131,52	2,63
4,875	20	39	2901,22	50149,82	40,79	107,39	122,51	391,983

Table 4.23: Discrepances between reference solution and IRBL estimation. Bangbae subway station train floor.

Image n.	ΔX **[m]**	ΔY **[m]**	ΔZ **[m]**	$\Delta\omega$ **[gon]**	$\Delta\varphi$ **[gon]**	$\Delta\kappa$ **[gon]**
11	-0,271	0,433	0,062	0,51	1,60	0,33
12	6,313	1,448	0,118	-5,43	1,33	-8,37
13	-1,186	1,03	0,192	-18,99	5,34	-18,06
14	1,084	0,23	-0,048	-6,32	1,11	5,44
15	-0,912	0,492	0,555	12,93	3,09	13,70
16	-1,526	-0,715	1,272	18,53	9,99	-19,27
17	-0,263	-0,164	0,008	-0,33	-0,65	-0,37
18	-0,734	-0,161	-0,05	11,30	0,62	10,76
19	0,592	0,269	-0,053	-0,82	-0,45	0,41
20	0,781	0,418	0,001	-1,00	0,35	0,50

Table 4.24: Statistical parameters results. Bangbae subway station train floor.

	ΔX **[m]**	ΔY **[m]**	ΔZ **[m]**	$\Delta\omega$ **[gon]**	$\Delta\varphi$ **[gon]**	$\Delta\kappa$ **[gon]**
min	-1,526	-0,715	-0,053	-18,99	-0,65	-19,267
max	6,313	1,448	1,272	18,53	9,99	13,697
med	0,388	0,328	0,206	1,04	2,23	-1,492
dev.st	0,917	0,612	0,417	10,85	3,24	10,95

3. **ETRI building**

- The discrepancies in X are always lower than 1 m in absolute value (Table 4.25);
- The standard deviations of the discrepancies in X is about 34 cm (Table 4.27);
- The discrepancies in Y are always lower than 1.5 m in absolute value for 6 images but the errors are of about 3 m (gross errors) for images 1_01.jpg, 2_03.jpg and 4_01.jpg. The shape of corridors of ETRI Building (long and narrow) and the poor geometry distribution of feature points due to the man-made environment decrease the IRBL accuracy;
- The standard deviations of the discrepancies in Y is about 2.1 m but it decreases to approximately 88 cm when poor solutions are not considered;
- The discrepancies in Z are lower than 0.90 m in absolute value for 7 images, but the errors are very large (over 8 m) for 2 images (3_05.jpg, 4_01.jpg): these mistakes are produced by erroneous extraction of reference images from DB because of the similarity of various floors in the ETRI building;
- The standard deviations of the discrepancies in Z, excluding gross errors, are of about 42 cm;
- The angular values are estimated with a precision of about 4 gon, but there is an image (2_03.jpg) where the angular external orientation parameter have a very poor solutions;
- There is an image (1_02.jpg) where it is not possible to obtain a solution according to the incorrect matching from reference image and smartphone image;
- The estimated averages are not significant for all the parameter, than there are not any systematic error estimation: precision and accuracy are substantially coincident.

Table 4.25: IRBL estimation for Bangbae subway station train floor.

Time [s]	**Image**	**n. points**	**X [m]**	**Y [m]**	**Z [m]**	ω [gon]	φ [gon]	κ [gon]
4,953	1	23	-71,648	-14,740	109,688	-62,857	104,906	163,858
4,755	2	12	No Solution					
4,781	3	12	-61,468	-21,254	115,179	365,816	-92,020	275,560
4,985	4	13	-61,005	-0,775	117,392	106,908	-0,668	-0,372
4,703	5	21	-61,426	-11,196	118,875	94,065	4,023	3,759
4,781	6	21	-61,495	4,457	101,737	-95,855	30,306	196,172
4,922	7	17	-61,641	-8,106	118,266	-100,077	33,273	197,913
4,890	8	24	-61,568	4,183	118,566	99,710	-31,692	4,989
4,781	9	18	-61,566	4,473	118,732	-96,552	31,133	198,673
4,875	10	27	-61,977	1,817	118,347	92,559	-46,902	-3,911

Table 4.26: Discrepances between reference solution and IRBL estimation. Bangbae subway station train floor.

n.	**Image**	ΔX **[m]**	ΔY **[m]**	ΔZ **[m]**	$\Delta\omega$ **[gon]**	$\Delta\varphi$ **[gon]**	$\Delta\kappa$ **[gon]**
1	1_01.jpg	-0,640	-2,563	0,561	2,666	0,737	-3,267
2	1_02.jpg	No Solution					
3	2_03.jpg	-0,233	3,736	-0,812	-55,750	12,179	-64,385
4	3_03.jpg	-0,839	-1,108	-0,006	-9,414	2,882	0,359
5	3_05.jpg	-0,343	0,726	-8,679	0,297	2,283	-1,443
6	4_01.jpg	-0,066	-3,578	8,482	-0,684	1,634	4,446
7	4_03.jpg	-0,320	-0,974	0,101	0,093	-1,015	1,298
8	5_01.jpg	-0,171	0,219	0,042	-2,017	-1,705	-4,938
9	5_04.jpg	-0,109	0,067	-0,106	-1,035	-0,386	1,935
0	5_06.jpg	0,343	1,036	0,274	0,553	3,986	4,284

The results obtained are summarize with their statistical parameter in Table 4.27

Table 4.27: Statistical parameters results. ETRI building.

	ΔX **[m]**	ΔY **[m]**	ΔZ **[m]**	$\Delta\omega$ **[gon]**	$\Delta\varphi$ **[gon]**	$\Delta\kappa$ **[gon]**
min	-0,839	-3,578	-0,812	-9,414	-1,705	-4,938
max	0,343	3,736	0,561	2,666	3,986	4,446
med	-0,264	-0,271	0,008	-1,193	1,052	0,334
st.dev.	0,340	2,136	0,423	3,593	1,995	3,379

4.2.7.2 Reliability of the IRBL algorithm

In the proposed procedure, the Fundamental Matrix estimation which express the epipolar geometry between query image (smartphone) and reference image (RGB-D) used a robust estimation algorithm based on RANSAC which insert a certain variability in the final estimate. This is function of the number of features samples extracted by SIFT. For each images and each test site,the IRBL procedure has been run many times to define the reliability of the solution. The results (Table 4.28) shown that the average values of discrepancies between different launches are small, always less than 2 cm. It is possible to affirm that the procedure is not affected by estimation biases. The standard deviation of discrepancies between different launches are small, always less than 18 cm. These values are less than $\frac{1}{4}$ standard deviation of IRBL solutions, then they are not significant respect to the IRBL precision (Figure 4.57).

Table 4.28: Statistical parameters of launch repetitions of IRBL code

	X [m]	**Y [m]**	**Z [m]**
mean	-0,016	-0,005	-0,004
st.d.	0,171	0,157	0,082

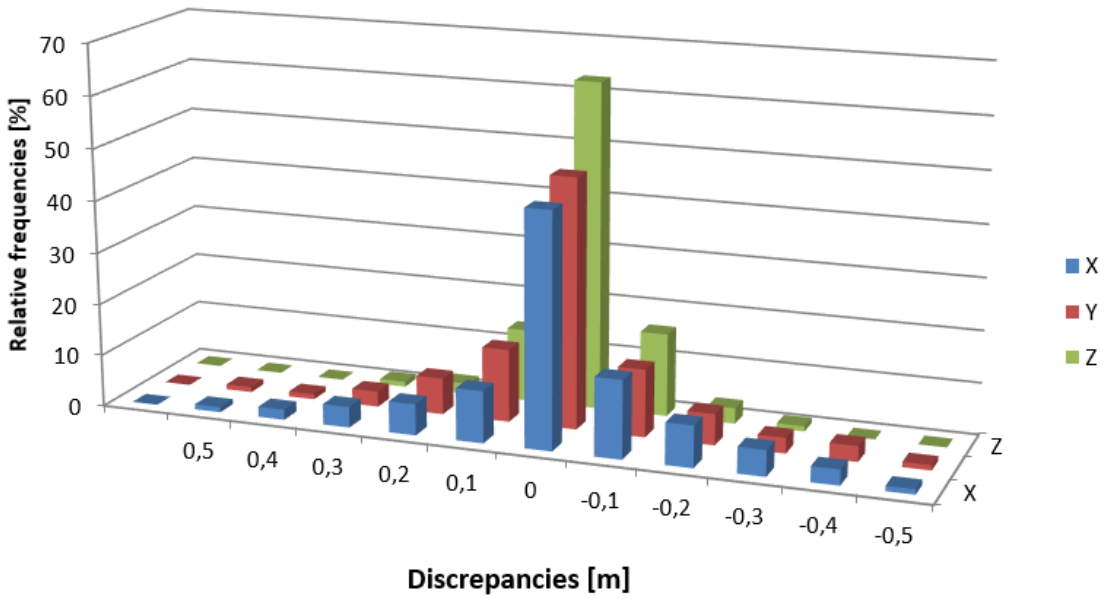


Figure 4.57: Distribution of estimation discrepancies for several runs of the IRBL algorithm.

4.2.7.3 Processing time

As stated before, the algorithm can be set in order to use SIFT or ORB feature extraction and matching algorithm. Although the pose estimation is similar, is interesting to observe the processing time required by the two procedure in relation with the different steps of IRBL. Table 4.29 shows the required time for each step of the workflow and the influence that they have on the overall time required. For SIFT-based procedure, IRBL requires about 5-6 seconds on a standard PC but the larger part (72%) is necessary to perform SIFT feature extraction. Looking at ORB-based algorithm, the total time is around 1.6 seconds with the 56% spent for the features extraction. These results shown the capability of the IRBL to perform in real time applications.

Table 4.29: IRBL processing time with SIFT and ORB algorithm.

	SIFT		**ORB**	
Steps	**Time [s]**	**%**	**Time [s]**	**%**
CDVS	0,875	16,512	0,875	54,82
Read/rescale	0,5	9,435	0,5	31,32
Sift / Orb	3,86	72,843	0,14	8,77
Matching	0,031	0,585	0,016	1
ratio test	0,001	0,018	0,001	0,06
Ransac	0,015	0,283	0,047	2,94
2Dto3D features	0,001	0,018	0,001	0,06
LevMarquard	0,016	0,301	0,016	1
total	5,299	100.000	1,596	100.00

4.2.8 Conclusions

The IRBL procedure is an image based indoor/outdoor positioning system based on the assumption that some a-priori information of the environment is provided. The advanteges of this procedure in urban application are numerous. First, the estimation can provide positioning both in absolute and local reference frame. Unlike numerous other positioning solution, the IRBL procedure provides not only position coordinates but also attitude of the camera with a good level of accuracy. These capability allows the use of these solution for Augmented Reality applications. The image based procedure is robust to noise produced by moving objects and people on the image view as demonstrated by the real case studies presented. On the other hand, the repetitiveness of some man-made environments could lead to an incorrect extraction of the reference image in the database. The presence of a database rich of information is a valuable features for other urban applications even if it requires a major effort in the survey of the environment and in the realization

of the 3D model. Looking at the results parameters of the proposed solution, a validation on accuracy, precision reliability and processing time has been made. The results shown meter level accuracy of the solution with sub-meter accuracy obtained excluding wrong correspondences. The precision is always under 70 cm. The reliability of the solution has been verified defining the nominal standard deviation of the IRBL procedure, as the same order of magnitude of the standard deviation obtained from ground truth comparison. The C++ implementation of the procedure requires 5 seconds to retrieve the position of the camera in real application. The validity of the proposed solution has been also demonstrated by the implementation of this C++ code in an application based service as shown in Figure 4.45.

4.3 GNSS smartphone positioning in distrisct scale environment

In Section 4.1 and 4.2 two positioning and navigation systems has been presented for two different urban spaces characterized by the scale of the environment. These two systems (UWB and IRBL) are mainly indoor systems which perform better in closed environment like office rooms or public buildings and infrastructures. In the following section of this work, the focus will move on outdoor positioning. Among the numerous systems the main technology used in this kind of situations is the GNSS. As seen in chapter 3.3, the miniaturization of the GNSS SoCs and their capability to perform accurately in adverse condition has increase the interest of the scientific community in smartphone positioning and navigation. The interest raise from the possibility to access directly to the raw measurements of these devices and to apply real-time and post processing algorithms used in geodetic applications or signal processing methods used in telecommunication. Previous research on smartphone positioning have demonstrated that, under good multipath condition, some meters accuracy is typical while in adverse condition it becomes around 10 meters. In general, the accuracy of the latest GNSS navigation messages is approximately 1 meters, making the main sources of errors, the front-end antenna and the receiver-dependent noise. For example Pesyna et al. in 2014 [85] deviates the analog signal acquired by the inner smartphone antenna directly to an external radio frequency front-end and GNSS receiver. This approach permitted to apply well-defined signal processing algorithm avoiding the limitation of the internal chipset. Whit this test a decrease of 11 dB in carrier to noise ratio with respect to a survey-grade antenna was observed. After the API level 24 implementation, first empirical test were focused on the evaluation of the smartphone raw observables. In [86] static measurements have been acquired with a Nexus 9 tablet and post-processed performing code phase single differences. The comparison with a reference solution have shown an error of around $\pm$ 20 meters, a pseudorange rate noise within $\pm$ 10 m/s and a

C/N0 average value of approximately 10 dB-Hz lower than geodetic receiver used in the same place. At the same time, advanced post-processing algorithms have been applied with the aim of achieving centimeter positioning with smartphones. To do this, the first step was the development of parsers capable of extracting the raw measurements in a format more suited to geodetic post-processing. In [87] the authors were able to reach sub-meter accuracy with on board doppler filtering algorithm and SBAS augmentation. In [88] decimeter level accuracy positioning has been achieved using post processing relative positioning i.e. by double differencing raw smartphones observation with those coming from a GNSS Continuously Operating Reference Stations (CORS) network.

In this section, a similar approach has been made in order to evaluate the performances of two commercial smartphones with an embedded multi-constellation GNSS chipset. In particular, the interest has been pointed on the challenges of urban environments. Multipath effects in urban canyons, voluntary and unintentionally interferences in crowded areas, no satellite visibility in shielded spaces.

4.3.1 Experimental setup

To consider all the previous aspect the proposed test consist in the data acquisition of GNSS measurements from different new generation smartphones on a known georeferenced point in Politecnico di Torino. The quality of these data has been analysed. Thanks to the knowledge of the reference point, different positioning procedures has been validated. In order to do this, as the raw data are provided in a no-standard format, a RINEX parser has been developed in matlab following the procedures already presented in Section 3.3.3. The different post-processing results has been compared and validated.

The smarpthone hardware used in the present application are two new generation devices whit installed the Android Nougat 7.0 Operating System (OS): Huawei P10plus and Samsung Galaxy S8+ (Figure 4.58). All the devices have a single frequency multi-constellation GNSS chipset installed, the Broadcom Limited Galileo-enabled BCM4774 GNSS chipset, able to collect GPS, GLONASS, Galileo and BEIDOU satellites data. No technical specifications are provided by the producer companies in order to define the exact location of the antenna in the smartphone body. Table 4.30 shows the main characteristics of the tested tools.

Figure 4.58: On the left Huawei P10 plus smarphone. On the right Samsung Galaxy S8+ smartphone.

Table 4.30: Specification of the tested smartphones.

Name	**Huawei P10 plus**	**Samsung Galaxy S8+**
OS	Andorid 7.0	Android 7.0
Chipset	Exynos 8895 Octa - EMEA	HiSilicon Kirin 960
CPU	Octa-core (4x2.3 - 4x1.7 GHz)	Octa-core (4x2.7G - 4x1.8 GHz)
GPU	Mali-G71 MP20	Mali-G71 MP8
GNSS	A-GPS, GLONASS, BDS, Galileo	A-GPS, GLONASS, BDS, Galileo
Sensors	Accelerometer, gyro, proximity, compass, barometer, heart rate, SpO2	Accelerometer, gyro, proximity, compass

Quality of the measurements

The first test consisted in acquiring 20 minutes of GNSS measurements with the two smartphones (same observation period) and observing the quality of the observation and the accuracy of the positioning obtained with a stand-alone post-processing. In order to do this, the smartphones has been located outdoor, near a fixed point with known absolute coordinate position. Having the real coordinates is possible to characterize the accuracy and precision of the smartphone relying only on the satellites measurements of the GNSS chipset. Figures 4.59 and 4.60 shows the pseudorange measurements for all the visible satellites and their orbital variation during the time acquisition. The number of visible satellites change between the different smartphones, in particuar 9 for Huawei P10 plus and 8 for Samsung Galaxy s8+.

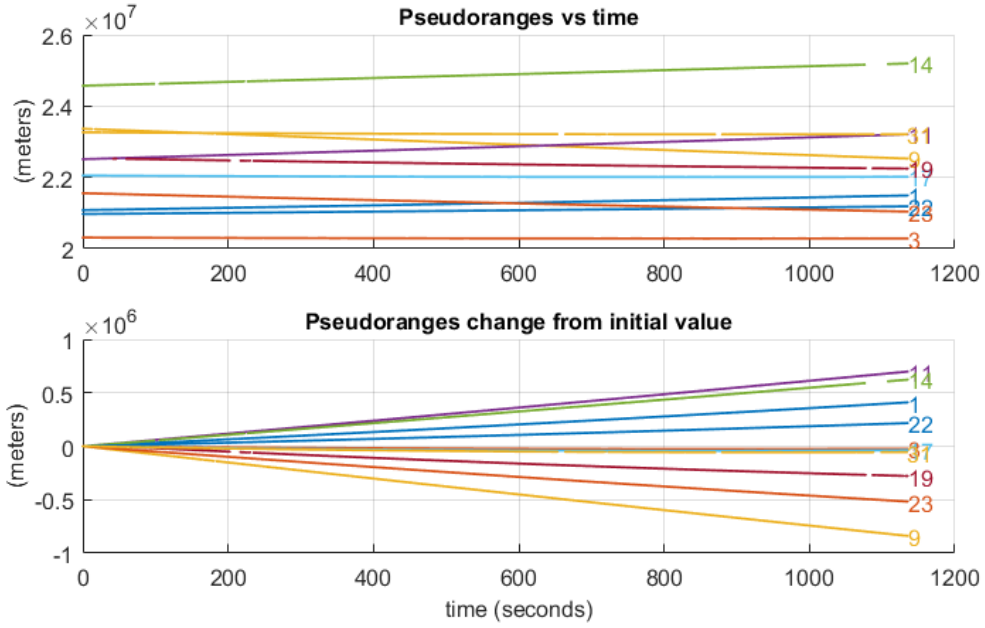


Figure 4.59: Raw pseudorange measurements acquired by Huawei P10plus smartphone.

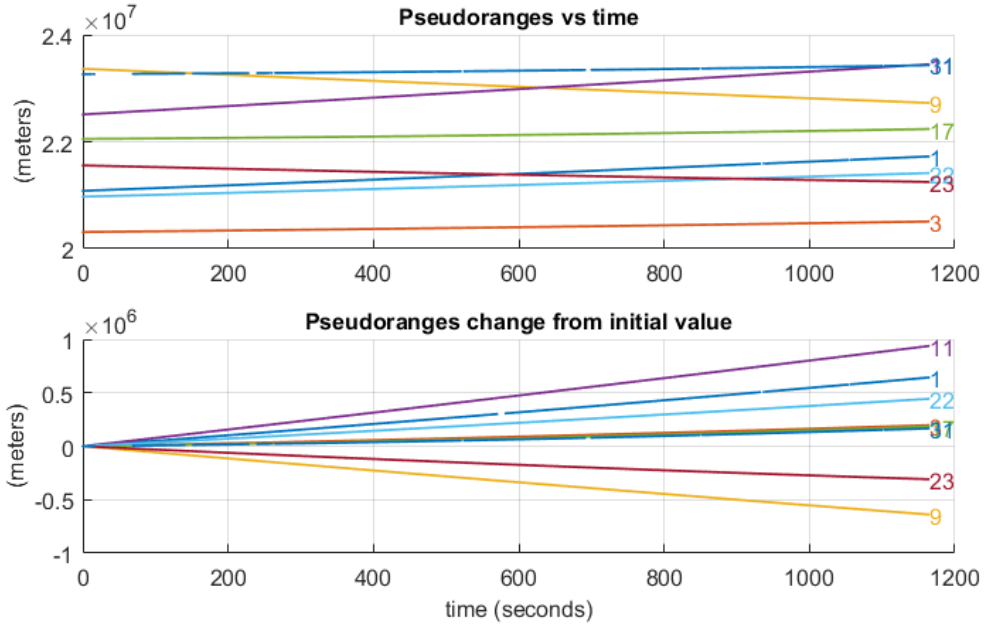


Figure 4.60: Raw pseudorange measurements acquired by Samsung S8+ smartphone.

On the same dataset, the Signal to Noise Ratio (SNR) has been plotted to verify

the quality of the signal. From Figure 4.61 and 4.62 is possible to note that some satellites exceed the limit under which the measurement is too noisy to be used (25 dB-Hz). Moreover the C/N0 raise over 40 dB-Hz rarely.

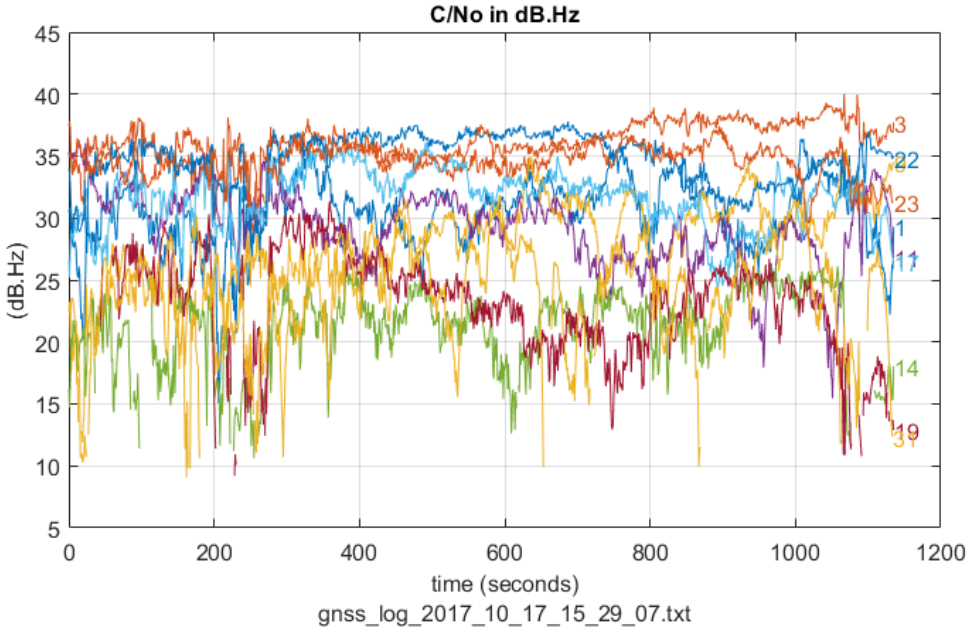


Figure 4.61: Signal to Noise Ratio (C/N0) of Huawei P10plus smartphone.

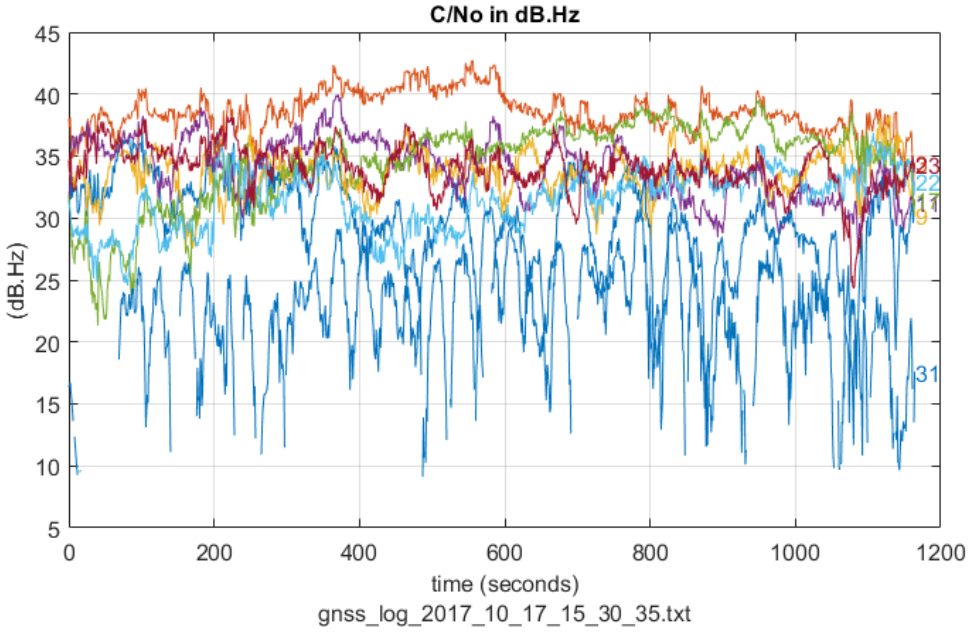


Figure 4.62: Signal to Noise Ratio (C/N0) of Samsung S8+ smartphone.

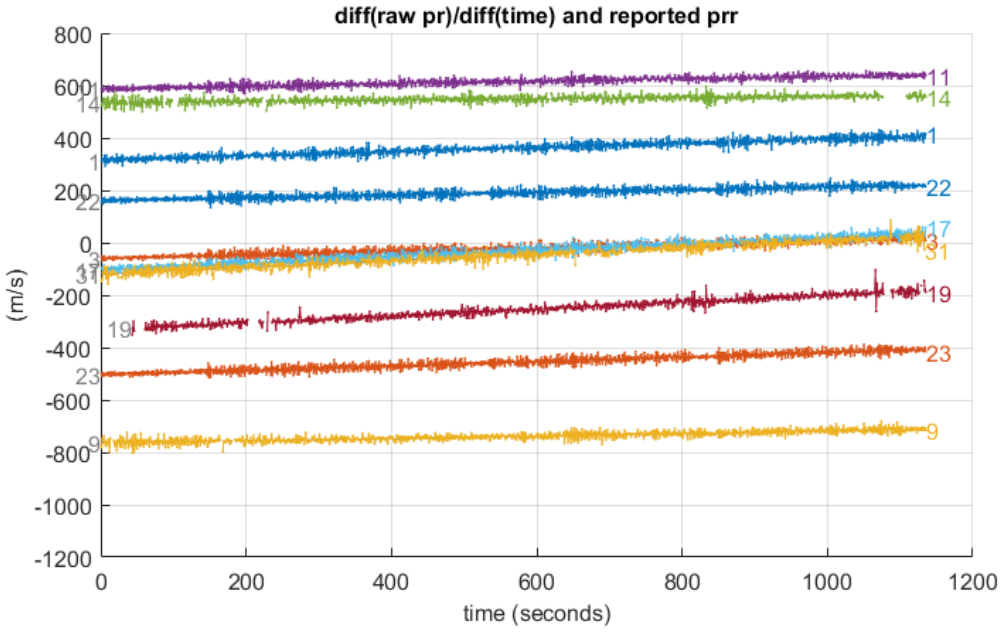


Figure 4.63: Differential pseudorange measurements acquired by Huawei P10 plus smartphone.

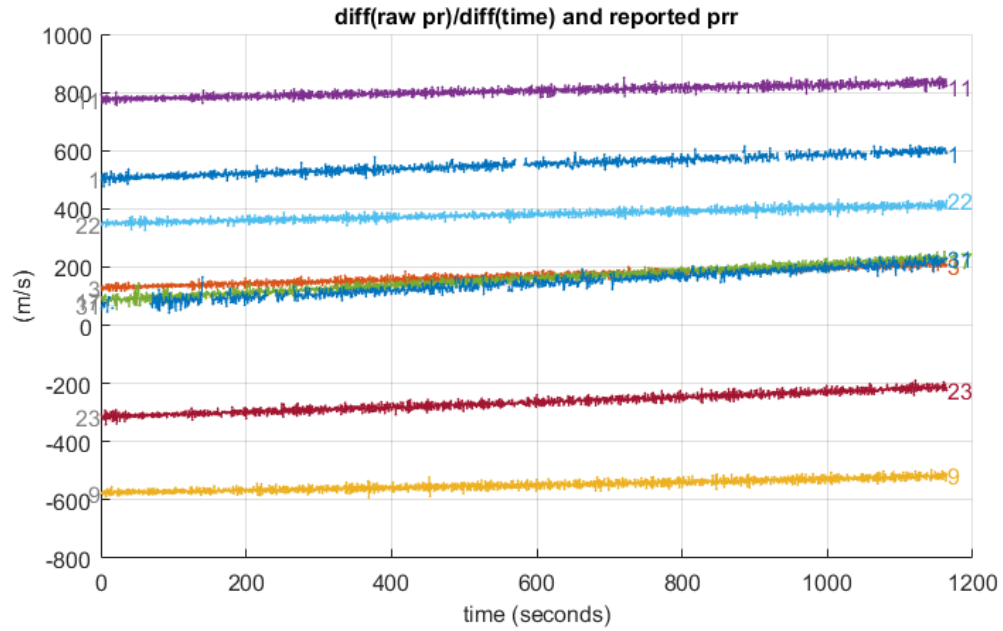


Figure 4.64: Differential pseudorange measurements acquired by Samsung S8+ smartphone.

Finally, the acquired pseudorange measurements has been used in an iterative

estimation procedure in order to validate the stand-alone position. In order to do this, a non linear weighted least square estimation (Equation 2.3) has been implemented and the obtained results are reported in Figure 4.65 for Huawei P10 plus smartphone and in Figure 4.66 for Samsung Galaxy S8+. For both the obtained values have been validated with the real position of the point. Considering the median value of the pose estimated, the 2D error for Huawei P10 plus is 2,7 meters with around the 50% of the points with 8,1 meters deviation. For the Samsung smartphone the 2D error is 4 meters with 50% of the point having 9.6 meters deviation. The accuracy and precision of the solution is reported in Figure 4.67 and 4.68 where is possible to observe for both devices a bias on the estimation along the Z component.

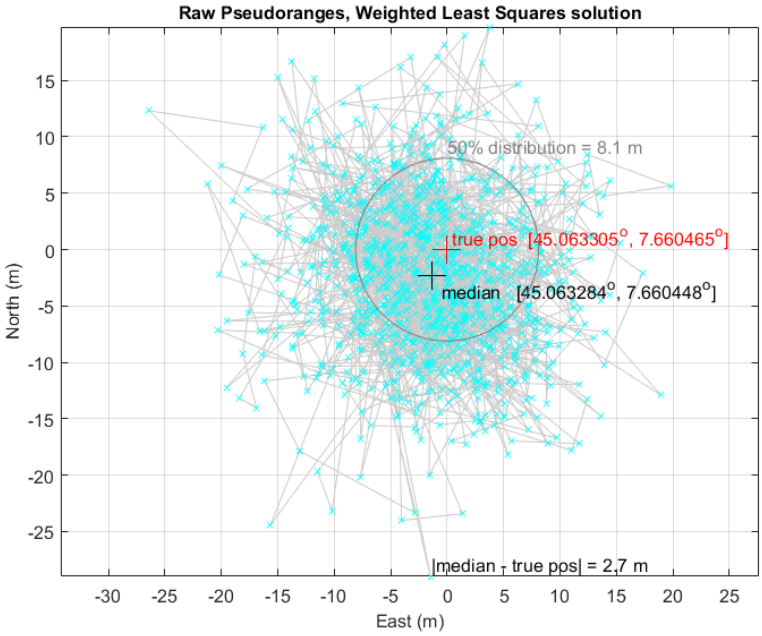


Figure 4.65: Weighted Least Square Positioning estimation from Huawei P10plus smartphone.

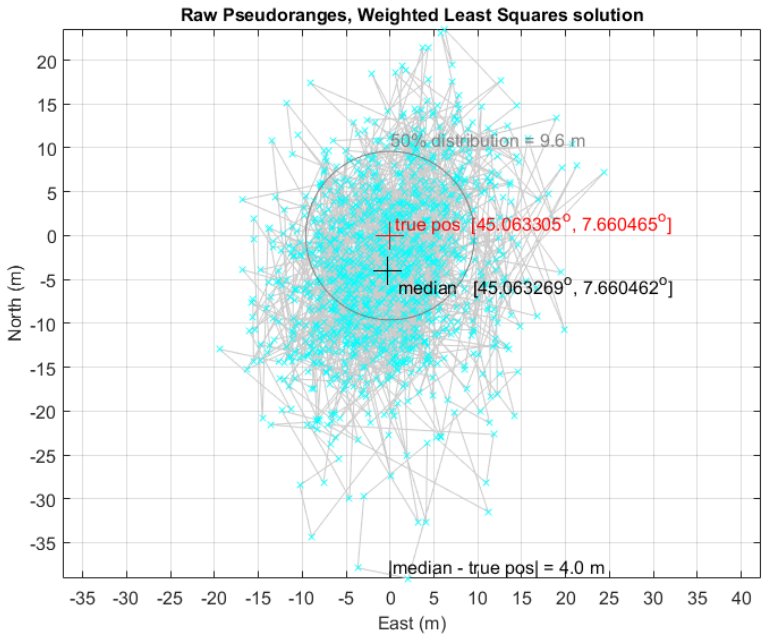


Figure 4.66: Weighted Least Square Positioning estimation from Samsung S8+ smartphone.

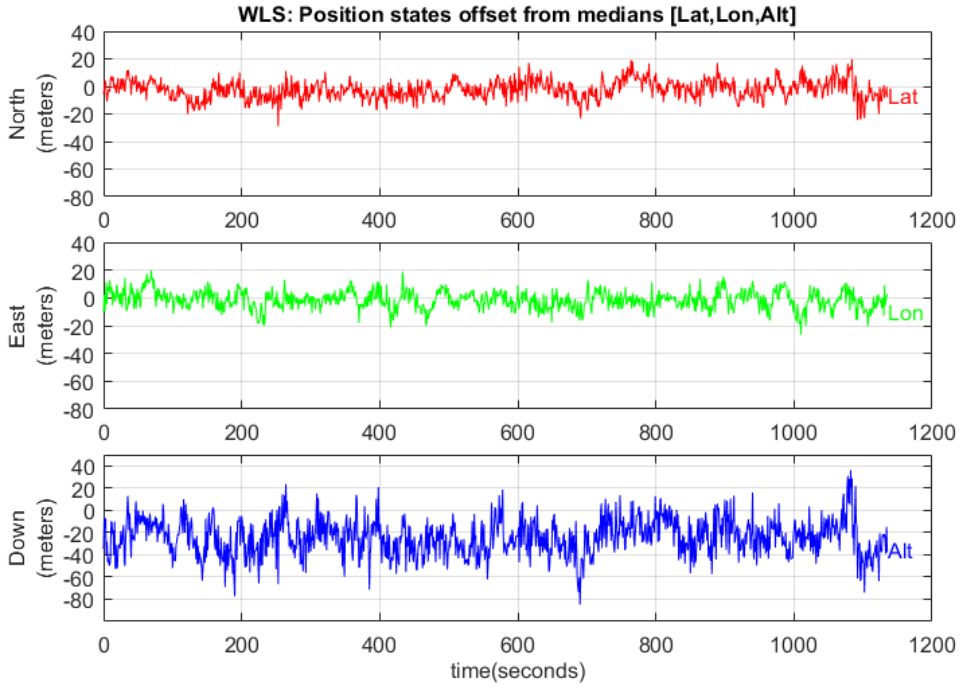


Figure 4.67: Three components WLS estimation from Huawei P10 plus smartphone.

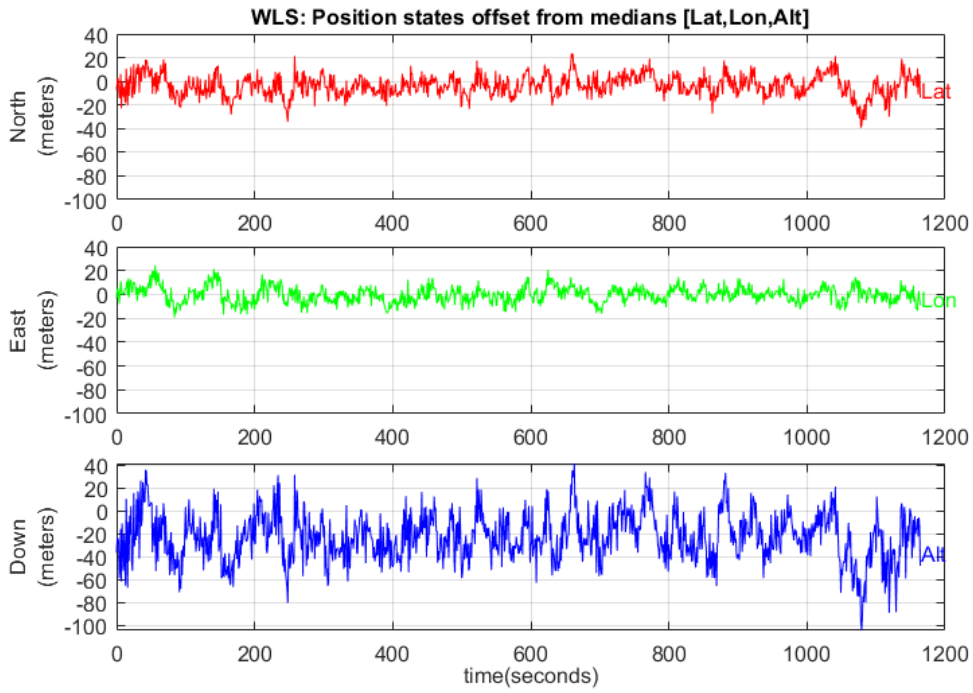


Figure 4.68: Three components WLS estimation from Samsung S8+ smartphone.

NRTK positioning

Among the numerous real-time and post-processing procedure to increase the accuracy in positioning of the smartphone devices, the NRTK technique is certainly one of the most interesting as permit to correct the stand-alone estimation with the use of aiding informations. In fact, the RTK network is a network of GNSS permanent stations spread all over the territory whose data are used to generate real-time correction for any rover receiver located inside the working area. With geodetic instruments, NRTK positioning allows high accuracy and precision performances in real time. Today, NRTK operates in several countries, such as Germany, Spain, England, Italy, China, some areas of the United States, Australia, and so on. Networks can have different extensions from small local networks with a mean inter-station distance of about 40-50 km to networks covering entire countries with mean inter-station distances of about 100-150 km. Generally, the network infrastructure consists of three segments. The first is composed of the so-called GNSS Continuously Operating Reference Station (CORS) networks, well-known located receivers spread across the territory to generate a distributed web. The second segment is the control centre, which collects and processes the data captured by the CORSs and broadcasts the differential corrections and saves the raw data for post-processing activities. Specifically, it fixes the ambiguity of all satellites for each permanent station and calculates ionospheric and tropospheric delays. Through

different interpolation models, it can provide calculated corrections to every point within the network. Such corrections can be sent in real time or can be used to create a virtual RINEX for a post-processing approach. The third part of this system is composed of the products generated by the control centre that can be sent to the users that rely on the service. The users, after a subscription, obtains RTK corrections that can be generated by several methods:

- Virtual reference station (VRS);
- Multi-reference station (MRS);
- Master-auxiliary corrections (MAX or MAC);
- Flächen-Korrektur-Parameter (FKP);
- Nearest station (NRT).
- DGPS/DGNSS

Among them,considering carrier phase measurements, only the VRS and the NRT services allow performing NRTK positioning with single-frequency receivers.

The idea on which the following test is based is to simulate a NRTK positioning using the smartphone measurements and the stream of RTK differential corrections acquired by an external device. To this purpose an u-blox LEA-M8T GNSS receiver coupled with a Gramin GA-38 antenna has been used and located in the fixed point near the smartphones. The u-blox receiver was setted to provide also raw GNSS measurements on L1 frequency. Using the RTKLIB v.2.4.3 b29 open-source software, NRTK solution considering VRS differential correction and stream of RTK differential correction have been acquired. The GNSS Raw measurements have been collecterd for 10 minutes considering a sampling rate of 1 Hz. Figure 4.69 shows the setup of the test. The image shows the different location of the smarpthone with respect the antenna position. For the result analysis, this lever arm is considered both for horizontal (23.6 cm) and up (13.7 cm) components (considering the mean position between two smartphones). Thus, the results shown in the following section refer to the same point, that is, where the Garmin antenna is installed.

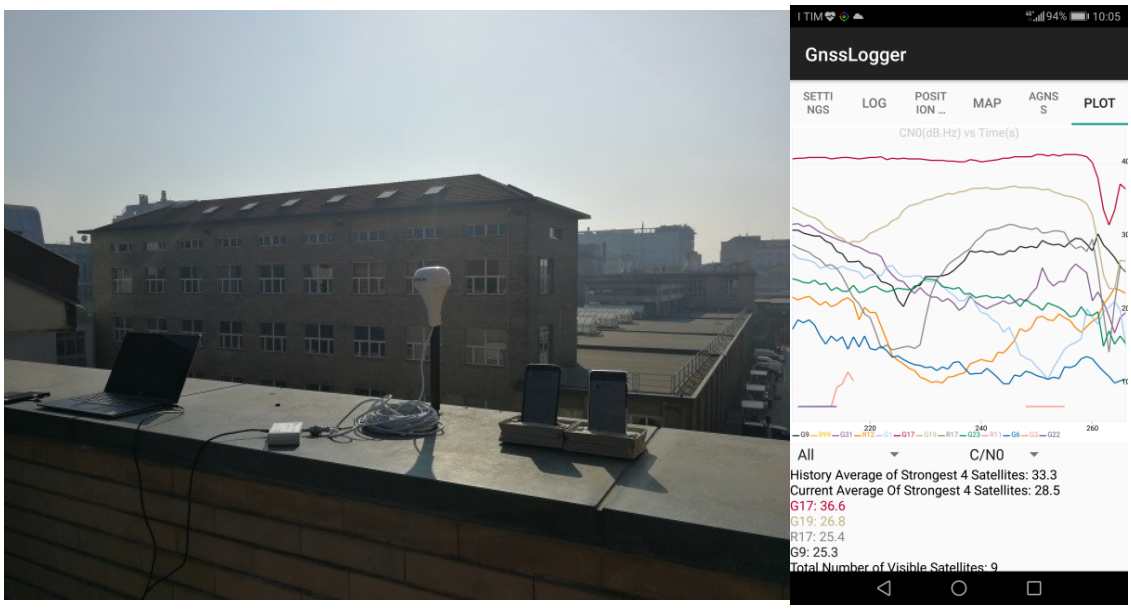


Figure 4.69: Rover test site: on the left, the external GNSS Garmin GA38 antenna installed on the roof of a building with the two different smartphones considered, while on the right, the smartphone app is running.

For these tests, the Servizio di Posizionamento Interregionale GNSS Piemonte – Lombardia (SPIN) GNSS CORS network [89] was considered. This network is managed by Leica GNSS Spider software and allows obtaining differential corrections for a user through the network transport of RTCM via Internet protocol (NTRIP) authentication after a free registration. This network, as shown in Figure 4.70, has a mean inter-station distance of about 50 km and is used for real-time and post-processing applications. The coordinates of all stations are obtained from a network adjustment computed with the Bernese GPS 5.0 software in the ETRF2000 reference frame.

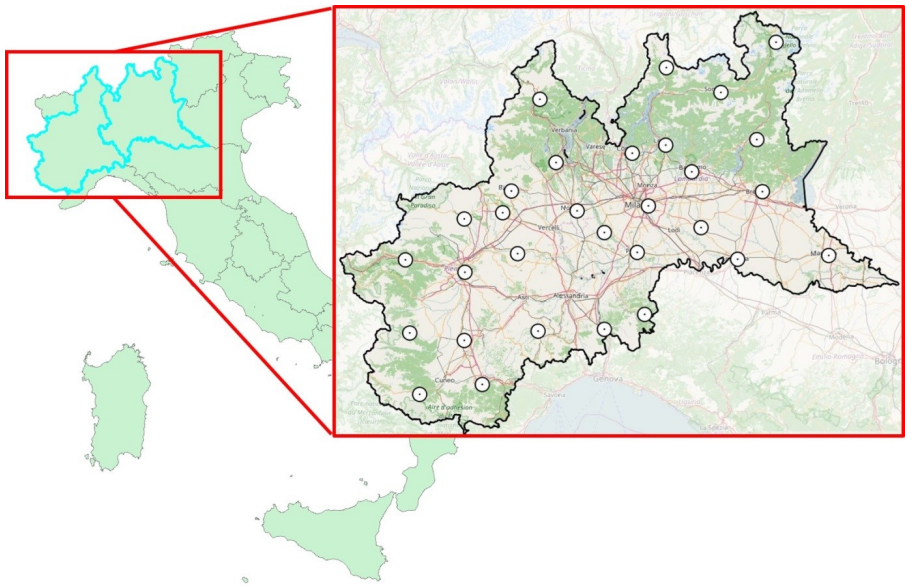


Figure 4.70: The SPIN GNSS network used for NRTK positioning.

In order to compare the performances of smartphones GNSS receivers with those obtainable with the u-blox, the application used for smartphone positioning considers the same algorithm for the ambiguity resolution that is available inside the RTKLIB software. This algorithm is based on the on-the-fly (OTF) integer ambiguity resolution method, where the values of integer ambiguities are obtained by solving an ILS (integer least square) problem thanks to a well-known efficient search strategy LAMBDA [90] and its extension MLAMBDA [91]. Moreover, a ratio factor of "ratio test" for standard integer ambiguity validation strategy has been considered. This factor, that can be considered also as threshold, means the ratio of the squared sum of the residuals with the second best integer vector to with the best integer vector. So, when the inequality $\frac{(\sigma_0^2)_{2nd}}{(\sigma_0^2)_{1st}}$ is satisfied, the ambiguities are defined as integer values, so it is possible to define that solution as "FIX", otherwise as "FLOAT". For these first experiments, the threshold value is set equal to 3.

The results of the NRTK positioning are reported in Figure 4.71 and Figure 4.72 where the trend of the difference between NRTK and reference solution is shown.The green points refer to solutions in which the phase ambiguities can be declared as 'fixed'. In these cases, the differences with respect to the reference coordinates are greater than the 'float' solutions. When the algorithm tries to fix the phase ambiguities, the solution becomes worse. This is maybe due to the quality of the smartphone measurements, that are more noisier than those obtainable with the u-blox receiver. Indeed, the software encounter some difficulties to fix the phase ambiguities in a correct way, and this is also confirmed if the time series analysis of the ratio value is analysed: as shown in Figure 4.73, it is possible to see both the

high variability in time of this estimation and the few cases where the ratio value is greater than the threshold if compared to those obtained with the u-blox receiver.

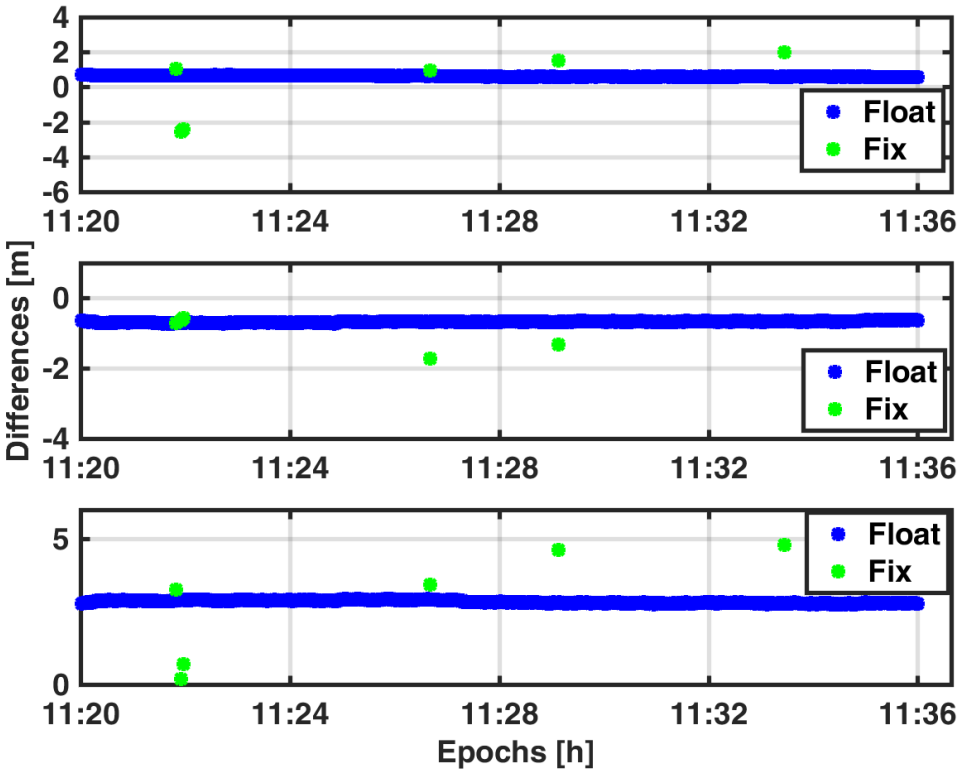


Figure 4.71: Positioning performances of P10 plus; from the top to the bottom of the figure represents the behaviour of the East, North, and Up components with respect to the reference values.

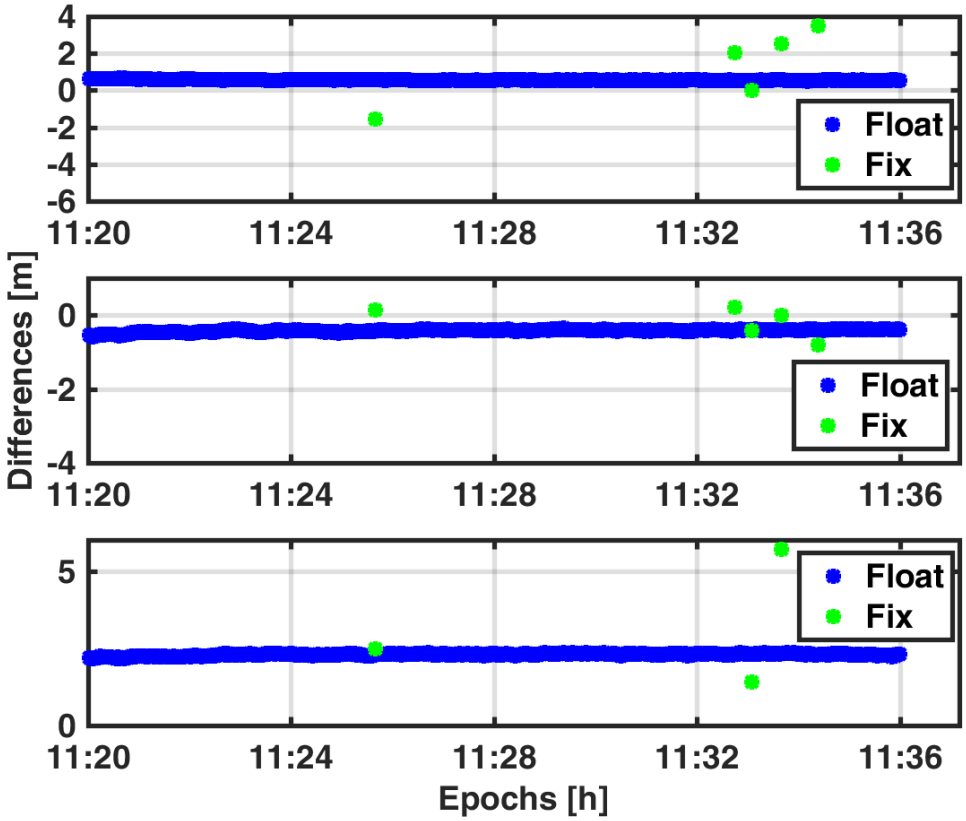


Figure 4.72: Positioning performances of Samsung Galaxy S8+; from the top to the bottom of the figure represents the behaviour of the East, North, and Up components with respect to the reference values.

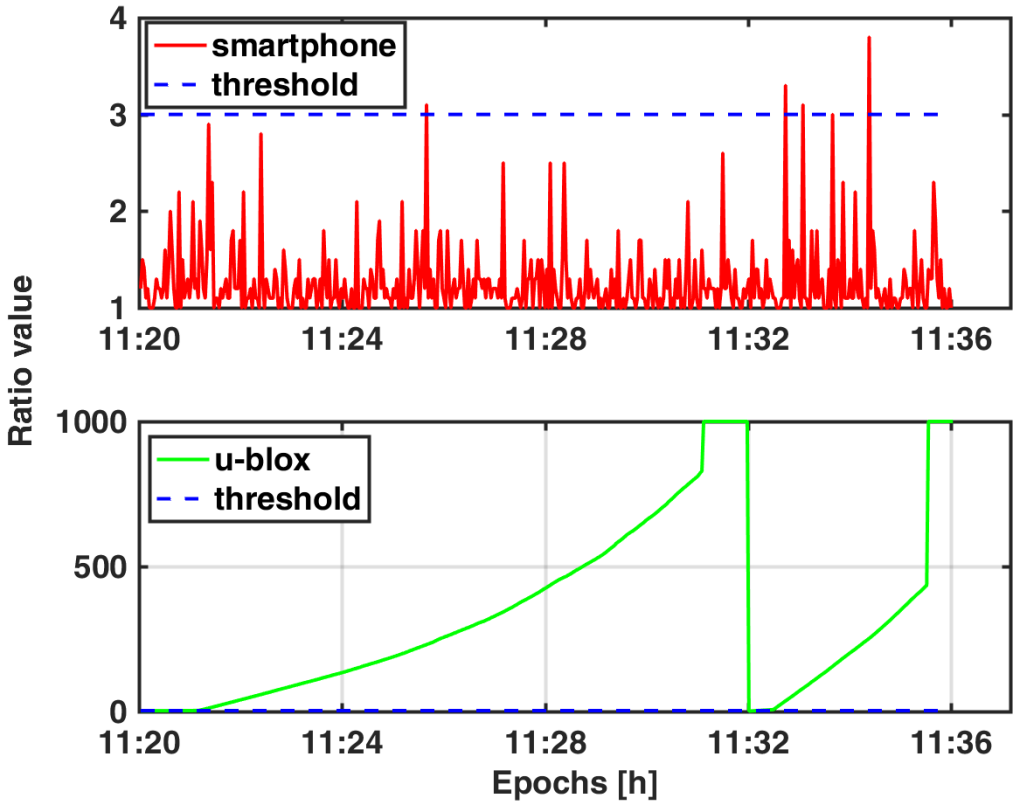


Figure 4.73: Time series analysis of the ratio value considering smartphone and u-blox receivers.

Increasing the threshold value, the percentage of the 'fix' solution decreases close to zero, but the quality in terms of differences between estimated and reference coordinates increases. In Table 4.31, the most significant statistical parameters are summarised.

Table 4.31: Statistical parameters related to the differences between estimated (NRTK) and reference coordinates considering GPS and GLONASS constellations.

	Samsung S8+		Huawei P10 plus	
	Mean [m]	Std [m]	Mean [m]	Std [m]
East	-0.200	0.078	-0.741	0.043
North	0.923	0.061	1.517	0.040
up	2.336	0.092	2.604	0.071

Table 4.31 shows that the two different smartphones provide quite different results. If the Samsung Galaxy S8+ gives a mean 2D value less than a meter, the differences related to the P10 plus are over 1.50 m. In both cases, the standard

deviations are about few centimetres; therefore, there are no gross errors and the solution is precise even if inaccurate.

Table 4.32: Statistical parameters related to the differences between estimated (NRTK) and reference coordinates, considering only the GPS constellation.

	Samsung S8+		Huawei P10 plus	
	Mean [m]	Std [m]	Mean [m]	Std [m]
East	-0.260	0.037	-0.686	0.023
North	0.532	0.026	0.616	0.036
up	2.452	0.046	2.853	0.052

Table 4.33: Statistical parameters related to the u-blox NRTK solutions with respect the reference coordinates.

	u-blox	
	Mean [m]	Std [m]
East	0.002	0.004
North	-0.001	0.004
Up	0.009	0.006

Considering the GPS-only solution, it is possible to improve the accuracy and precision. Table 4.32 shows that both smartphones obtain a 2D accuracy around 60 cm with a standard deviation of a couple of centimetres. Thus, for these kinds of receivers, the multi-constellation approach does not provide any benefit if NRTK positioning is computed. Comparing these results with those obtainable with the u-blox receiver and Garmin antenna, it is possible to note a completely different behaviour in terms of accuracy. The u-blox provides excellent results, comparable to those available in the literature as seen in Table 4.33.

4.3.2 Conclusions

From this study, it is confirmed that it is possible to perform NRTK positioning with smartphones. It is not so easy to reach an accuracy of a few centimetres because in addition to some problems such as multipath and imaging effects, one of the main issues is still to know where the GNSS antenna is inside the smartphone. In most cases, this information is unavailable. Thus, the real problem is to know the exact position of the smartphone antenna. While the chipset position is quite well represented in the manufacturer schemas, the antenna position is usually not highlighted. Therefore some assumptions should be made. One could be to approximate the position on the centre of the smartphone. The assumption made in

this work is to consider the size of the smartphone as the tolerance for the precision of positioning. The phase-centre identification of the smartphone GNSS antenna will be a subject of other research. In this context, it does not make sense to try to fix the phase ambiguities. It is better to have a good 'float' solution rather than a bad 'fix' solution. Two different smartphones, with different internal chipsets have been tested in a CORS network with a mean inter-station distance of about 50 km, considering both VRS and nearest corrections. The results have shown satisfactory performance in terms of precision but not from the accuracy perspective. Even if the two smartphones provide slightly different results, the accuracy obtainable today is greater than 1 m with a precision of few centimetres, especially if only the GPS constellation is considered. These results will open new frontiers in terms of real-time positioning with portable devices, especially for rapid mapping or emergency situations. In the future, it will be interesting to test single-base RTK positioning, considering a mass-market master station, to analyse what happens if the rover user is located where the NRTK positioning is not available due to the lack of a CORS network.

Chapter 5

UWB/INS/VISION HYBRIDIZATION FOR UNDERGROUND NAVIGATION

In this chapter a low-cost multisensor solution for pedestrian navigation is presented. The solution utilize as primary sensor technologies the ones presented in Chapter 3: Imaging sensors, UWB radio communication systems, GNSS smartphone positioning and inertial navigation systems. The reasons why these sensors have been selected are related with the challenges of the indoor navigation and will be discussed later in this Chapter. The solution proposed is the result of a research project named "Kaivos project" which has the aims to navigate in an extreme challenging environment: an underground mine. For this reason, this chapter will start with the description of the project and the objectives of the same. Then, the issues of this critical environment and their relation with the choice of the sensors will be discussed. The chapter will continue with the description of the data acquisition in a real mines and the analysis of the collected data. After this, the proposed fusion algorithm, based on the Extended Kalman Filter, will be discussed and finally the validation of the system will be presented. The proposed approach is based on the fusion of different kind of data (ranges, pressure, features, estimated position, velocity and attitude) to exploit the advantages and benefits of each acquiring sensor and to limit their issues in stand-alone configuration.

5.1 The Kaivos Project

The action plan [92] of the Ministry of Employment and Economy for promoting sustainable extractive industry in Finland, aware of the importance of the availability of natural resources, aims ..." to assure sufficient raw material supply by more intensive exploration and mapping of new mineral resources and by recycling

of metals, mineral products and waste materials". Currently, during the waste material flow in mining sector, a significant amount of valuable precious and critical materials are lost. There is therefore a need for more efficient and rapid determination of mineralogical composition of excavated rocks, in order to optimize both the selective mining extraction and the reuse of geometallurgical elements instead of wasting [92]. To accomplish these requirements researchers have focused their effort in developing smart underground technologies exploiting several scientific sectors, from robotics, image processing, surveying to navigation and positioning. The research related to the mining industry, not only have the aims to enhance the profitability of mining companies improving the production efficiency but also to face with occupational safety, environmental protection and emergency management [93]. In this panorama, ones of the main state-of-art technologies used in mines are the laser scanner techniques which are currently used for acquire information relative to the hyperspectral response of the rocks related to their mineralogical composition [94]. This technology can provide a dense tridimensional model of the mine useful for spatial analysis, process management, volume computation and more. This spatial information can be generated by the light detection and ranging technology in absolute reference system if the accurate position of the LiDAR instruments is known. For this purpose, a positioning system should be integrated in the laser acquisition platform. Under these conditions, the Kaivos project has been funded with the aims of "developing active 3-dimensional hyperspectral imaging combined with multi-sensor positioning and integrated in autonomous platforms in order to automatic identify and mapping minerals for efficiency and safety of the mining industry" [95]. The investigation related to Hyperspectral LiDAR (HSL) developed by the SINA research group of the Finnish Geospatial Research Institute (FGI HSL) has been already presented in [96] and [97] as well as the work in progress related to the multi-sensor positioning [98]. Staring from the already implemented procedure based on an INS/GNSS integration, this work will present a new framework that could provide both accurate positioning in underground mines to the FGI HSL and seamless navigation solution of vehicles in the transition from an open space to an indoor mine. The advantages of the new solution with respect to the previous one will be the possibility to achieve centimeter accuracy in static condition also in a GNSS denied environment and the possibility to define a seamless solution for navigation with integration of different low cost sensors.

5.1.1 Challenges in underground mine

Accurate positioning of the laser sensors is essential for the usability of the technology. Unfortunately the systems based on GNSS, which are the most used in machinery navigation and tracking, are not available in indoor situations and underground environments. Therefore, the main challenge in these type of scenarios is provide location information using other sensors and methods rather than

GNSS. In the last years, the IPS research field has provided numerous solutions mostly based in integration or hybridization of different sensors in order to overcome the singular limitation of each stand-alone technology. These research have proven that the multi-sensor fusion is the right way to follow in order to provide the better possible solution for navigation. Inertial sensors, magnetometers, radiofrequency networks, imaging and vision sensors, ultrasounds and more has been used in different integration techniques and with different purpose. Which sensor should be selected depends mostly on the environment on which the system have to operate. Mines or underground environments in general are quite challenging as the surrounding can influence the sensors performances and consequently the solution estimation (Figure 5.1). In order to approaching the positioning problem in mines, the previous issues and how the sensor are affected has been analysed:

1. **Indoor environment.** As is known, in indoor spaces the GNSS signal can't reach the receivers and the satellite positioning can't be applied. Underground, the most used technologies are based on radiofrequency sensors spreads in the environment able to define position of a receiver with ToF, ToA, RSSI measurements and more. The sensors are usually based on Wi-Fi signals, infrared, ultrasound, Bluetooth and UWB. In this work UWB systems are chosen. As described in Section 3.1, UWB is a radio communication technology which thanks to the principle of the "two way time of flight" (TW-TOF) allows to measure multiple ranges between fixed anchor and mobiles tag and define the accurate position with trilateration algorithms. UWB systems, compared to other indoor navigation systems, have a high time resolution that allows to measure ranges with accuracy of some centimeters. Their signal characteristics make it also resistant to multipath and interference.

2. **Light condition.** Several papers have shown that positioning and navigation systems based on camera sensors and computer vision or photogrammetric algorithms can provide very accurate solution respect to other state of art technologies [99]. Most of the benefits of these kind of systems depends on the possibility to extract salient information from RGB digital images. These unique points on the images are the fundamental information for the subsequent positioning algorithm. Unfortunately, the detection of these features on the images is based on intensity and color information of the pixels. This means that in a dark environment, like a mine, it is difficult to extract sufficient features to apply the estimation procedures. Although these vision sensors seem to be not suitable for mines, they have the big advantages to provide attitude information. Therefore it has been chosen to test thermal and infrared cameras which functionality in dark environment has been already demonstrated [100] [101]. Using these sensors is possible to extract user motion in a vision-aided positioning systems with visual odometry algorithms.

3. **Magnetic effects.** Precise heading information is vital for various operations in underground mines. Getting absolute heading relative to some local or global reference can be difficult due to the circumstances and lack of available infrastructure. Local infrastructure dependent methods require sufficient amount of movement and consequently tend to suffer from low precision. Gyrocompassing techniques are not yet suitable for low-cost portable systems. Magnetometer based methods fulfill the precision and portability requirements, but suffer from distortions due to ferrometallic and magnetic objects in the vicinity of the system. The amount of magnetic field distortions can vary depending on the type of mine. For example, in ferrochrome rich mines the distortions can be significant enough to render magnetometer based heading unreliable. To alleviate the effect of these distortions, could be interesting to examine the use of multiple magnetometers in different configurations. The use of multiple magnetometers allows easier distinction of distorted field readings in addition to providing statistical improvements to sensor noise.

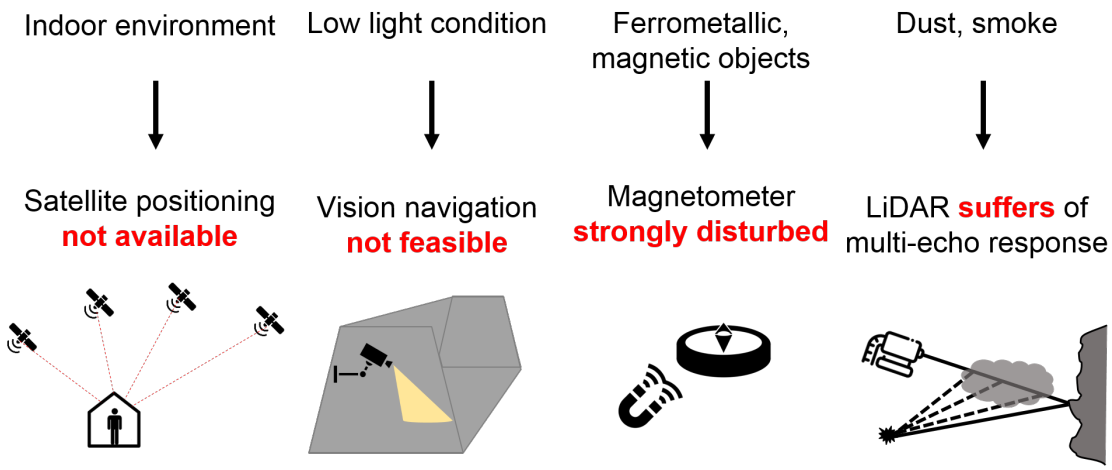


Figure 5.1: Measurements affected by the mine environment.

5.2 Proposed estimation framework

Starting from preliminary studies for a seamless positioning system based on a plethora of sensors, this work present a new framework for defining accurate positioning and vehicle navigation even in extreme environment such as underground mines. The main investigation fields composing the proposed solution are the multi-sensor navigation and positioning platform and the fusion algorithms with particular attention to the context awareness and the accuracy requirements. Let's imagine the FGI HSL mounted on a vehicles operation in a mine site. Assume that the task is to acquire spectral sign information from a perimeter wall of an open

pit. Firstly, the vehicle should navigate in the mine and reach a specific points. One arrived in the designated point, the absolute position of the HSL should be estimated. For both these tasks, in the case that sufficient number of satellites are visible, the standard GNSS procedure can be used to perform a complete position, velocity and timing (PVT) solution. If some obstacle or some short tunnel blocks the GNSS signal for few moments, an INS/GNSS integration can make the system immune to momentary GNSS outages. Now suppose that the HSL has to be moved to a specific point in an underground mine, moving from an open place to a closed tunnel. In this case,tThe GNSS/INS integration could provide a reliable navigation solution only for few minutes if a navigation grade inertial measurement unit (IMU) is used. This time will be even shorter if the IMU is a low cost Micro Electro-Mechanical System (MEMS) sensor. Moreover, once that the HSL is on the target point, there is no way to provide a precise position. The proposed framework can be perfectly suitable to face this typical situation in mining site as different sensors are selected to go towards a seamless solution integrating measurements adaptively based on availability and reliability of the different systems. Another interesting aspect of this framework, which is not object of the thesis, is the possibility to exploit RADAR capability of UWB sensors to implement a personnel tracking system. Implementing this feature in a hyperspectral LiDAR survey procedure could be very useful to increase the safety level of the operation (Figure 5.2).

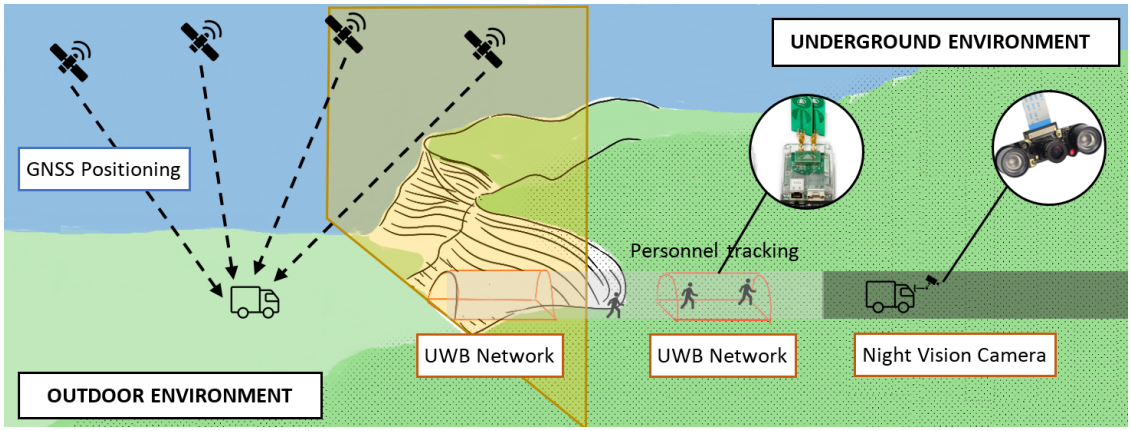


Figure 5.2: The proposed framework illustrated in a hypothetical scenario where means of transport and workers moves in a mining site both outdoor and in underground environment.

5.2.1 Low-cost multisensor platform description

In previous sections the operational framework related to the challenges of underground navigation has been presented. In this section the low-cost sensors chosen to be integrated in a single multisensor platform are described. Some has been

already presented in chapter 4 while others have been selected for the specific case study and will be briefly described. The measurements obtained from each sensor and how they are fused is deeply explained in section 5.2.2. The platform implementation has moved from a preliminary setup to a final one on which, data analysis, measurements characterization and positioning estimation validation has been made.

1. **GNSS.** With the purpose to provide a trajectory solution in favourable open sky environment a consumer grade single frequency receiver installed on a new-generation smartphone (i.e. BROADCOM GNSS chipset mounted on a Huawei P10) has been implemented in the platform 5.3. As demonstrated in Chapter 4 with this kind of low-cost System of Chipset (SOC) is possible to obtain a sub metrical accuracy in Near-Real Time Kinematic (RTK) navigation. Although the presence of the smartphone in the platform, no data have been acquired during the test performed in the underground mine. This due to the fact that there was no direct connection between the outdoor environment and the section of the tunnel on which the test has been made. Section 4.3 has given already an overview on the use of this technology in urban canyon and results for positioning. Further test in using this mass market devices for outdoor/indoor navigation are not object of this chapter and could be investigated in future research.

2. **INS.** The three-dimensional inertial sensors used for the test was an Xsense MTi-G-710 GNSS inertial sensor unit. It is composed by 3 axis accelerometer, 3 axis gyroscope measuring specific forces and rotation rates of a body on which the sensors is installed. There is also a 3 axis magnetometer and a baroaltimeter. These forces are used in INS mechanization to firstly deduce acceleration of the body and consequently to estimate position, velocity and attitude of the same. In literature, one of the most beneficial form of hybridization is the integration of GNSS with inertial sensors, also known as GNSS/INS integration. This integration allows to the GNSS system to provide attitude information otherwise not measured, can give more robustness to the navigation solution and can correct INS drift errors. The Xsense is able to cope with multi-constellation GNSS outages. The baroaltimeter estimates the altitude as function of the pressure. The limitation of barometer sensors is their noisy behavior due to rapid changes in pressure in the environment unrelated with the altitude. Table 5.1. shows the product characteristics as reported from the vendor [102].

Table 5.1: Sensor specification provided by the manufacturer.

Specification	**Gyroscopes**	**Accelerometers**
Standard full range	+/- 450 º/s*	+/- 20 g
Initial bias error	0.2 º/s	5 mg
In-run bias stability	10 º/h	$15\mu g$
Bandwidth (-3 dB)	415 Hz	375 Hz
Noise density	$0.01deg\sqrt{Hz}$	$60\mu g\sqrt{Hz}$
g-sensitivity (calibrated)	0.003 º/s/g	N/A
Non-orthogonality	0.05 deg	0.05 deg
Non-linearity	0.01%	0.1%
	Magnetometer	**Barometer**
Standard full range	+/- 8 G	300-1100 hPa
Total RMS noise	0.5 mG	3.6 Pa
Non-linearity	0.2%	N/A
Resolution	0.25 mG	8 cm (sea level, 15 °C)

3. **Ultra-wide band.** This radio technology is based on a network of transceivers able to send and receive very high temporal resolution and bandwidth signal from which is possible to estimate accurate range measurements (usually from 30 cm to 1 cm accurate). The two way time of flight measurement principle allows to overcome the synchronization problem between the clocks of the different devices as the time is always referred to the same receiver that is also the mobile tag of which the position have to be estimated. In the present work, the TimeDomain P440 PulsON solution has been used for accurate static positioning and high rate navigation. The characteristic of this sensor has been already presented in Chapter 4 (Table 4.1).

4. **Thermal Camera.** The commercial FliR One PRO camera has been selected in the first implementation of the platform (Figure 5.3). The aims of using a thermal camera is to apply classical visual odometry approach, usually made with RGB camera, in a dark environment, where image features are difficult to detect with the use of classical vision cameras. The FliR One PRO camera has two different imaging sensor, an RGB camera of 1,5 MPx and a passive thermal sensor with a resolution of 160x120 pixels. The pixel size of the thermal sensor is 12 micrometers while the spectral range is from 8 to 14 micrometers. The maximum frame rate of both sensor is 8.7Hz. Finally, the cost of this commercial sensor is 500 €.

5. **Infrared Camera.** The vision sensor used in the dark areas of the mines and implemented in the final version of the platform is a 5 Megapixel OV5647 Raspberry Pi Camera. It is composed by a sensor of 25mm x 24 mm with a

maximum resolution of 1080p. It can be used in daylight condition thanks to an embedded infrared-cut filter and in dark environment thanks to 2 attached infrared LEDs. The cost of this sensor is around 50$. Table 5.2 shows the sensor characteristics.

Table 5.2: OV5647 imaging sensor characteristics.

pixel size	1.4 micron x 1.4 micron
sensor size	1/4" equal to
focal length	3,6 mm
active array size	1592 x 1944 pix
image transfer rate	1080p: 30 fps

The experimental setup of the two platform are reported in Figure 5.3.

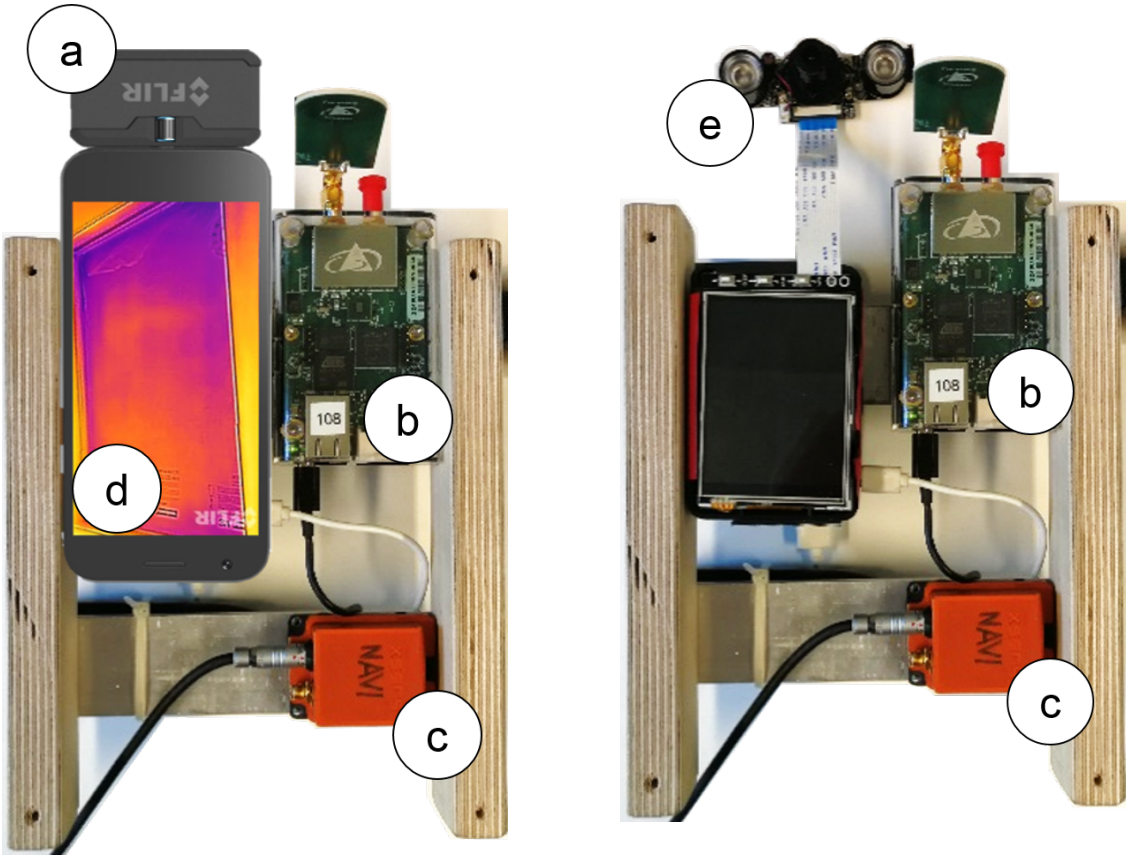


Figure 5.3: Multisensor platform setup 1: a) Thermal camera, b) UWB P440, c) Xsense IMU, d) smartphone with GNSS chipset. Mutisensor platform final setup: The thermal imaging camera and the smarthpone is substitute by the e) infrared camera sensor and a raspberry.

5.2.2 Fusion algorithm

The proposed methodology for underground navigation (Figure 5.4) is composed by two main cores: one based on the Computer Vision field, in particular the visual odometry navigation, and the other that follows the classical navigation algorithms, the hybridization of inertial and GNSS measurements. This two core are integrated to obtain an unique navigation solution providing positioning, velocity and attitude information of the platform previous presented. The currently proposed solution could be easily obtained with better performances from available commercial solution and previous researches which exploit high grade sensors with an high cost. When low-cost sensors are used, the feasibility of the navigation becomes more challenging and the quality of the estimated solution decrease. In this scenario, the proposed fusion algorithm takes into account all the issues of the underground navigation as well as the noise introduced by the low-cost sensors and propose a new estimation solution. In particular, for what concern the Computer Vision core, the main problems to face with are the low resolution of the imaging sensor and the low-light condition of the environment. The first problem generate blurring effect to the images acquired during the motion, the second make impossible the use of RGB cameras in mines. Regarding the GNSS/INS hybridization, is evident that in an underground mine, the GNSS receiver can't receive any satellite signal and so no output solution can be produced for the classical Kalman Filter integration. In this case the proposed methodology proposes to replace the GNSS positioning technology with an UWB-based positioning system. This allows to perform an UWB/INS hybridization with the same theoretical principles of the GNSS-based but aiding an important advantage: the seamlessness of the navigation in the transition from outdoor to indoor environment. In fact, ensuring an accurate georeferencing of the UWB network located in the mine, it is possible to switch from the GNSS/INS integration to the UWB/INS integration through a geofencing algorithm that recognize when the platform enters in the sensor network [103].

The procedure of the proposed algorithm is presented in Figure 5.4.

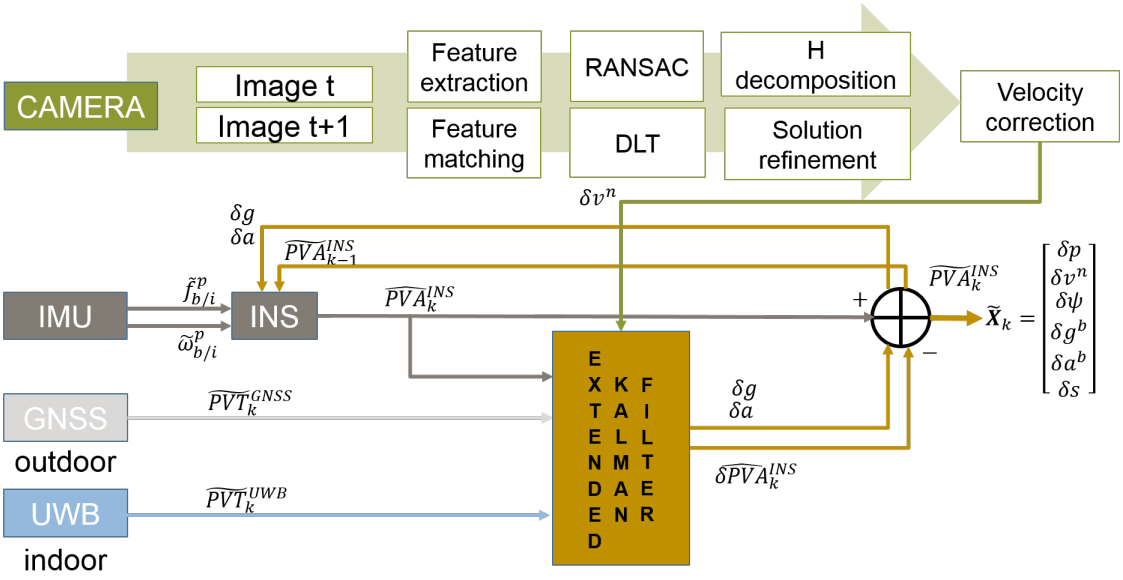


Figure 5.4: UWB/INS/VISION Hybridization algorithm.

The estimation state is composed by position (latitude, longitude, altitude), velocity vectors along the north, east and down axis, attitude (roll, pitch, heading), acceleration and gyroscopic bias. An EKF is used to perform prediction and update state estimation, the velocity vector provided by the UWB is augmented with the velocity estimation obtained by the visual odometry procedure. The filter is used in feedback form so that when a measurement is available from a sensor, the error is computed using the Kalman filter which is then used to correct the inertial sensor measurements and navigation parameters. On top of this algorithm is also possible to introduce relative altitude estimation from a barometer sensor and heading compensation from magnetic measurements.

5.2.2.1 GNSS/INS hybridization and UWB/INS Hybridization

The GNSS receiver and the MTi-G-710 GNSS inertial sensor unit are the systems that integrated in a loosely coupled algorithm represent the starting navigation solution for the open pit. The GNSS receiver provide measurements at 1 Hz, while the IMU works at 400 Hz. This high rate is fundamental to fill the gap between two subsequent GNSS measurements. The GNSS position and velocity are used to estimate the INS error. The barometer and the magnetometer integrated inside the Xsense sensor provides measurements respectively at 100 Hz and 40 Hz. When the systems moves underground the GNSS receiver is no more able to provide measurements so it has been substituted by the UWB receiver that enters in the UWB network of fixed anchor nodes. Also the UWB can work at 1Hz providing positioning in the mines. The previous loosely coupled integration is used in this case

simply using the UWB data to update the INS navigation. Theoretically, the algorithm follows the procedure presented in Section 3.4.4 and summarize in Fugure 5.5.

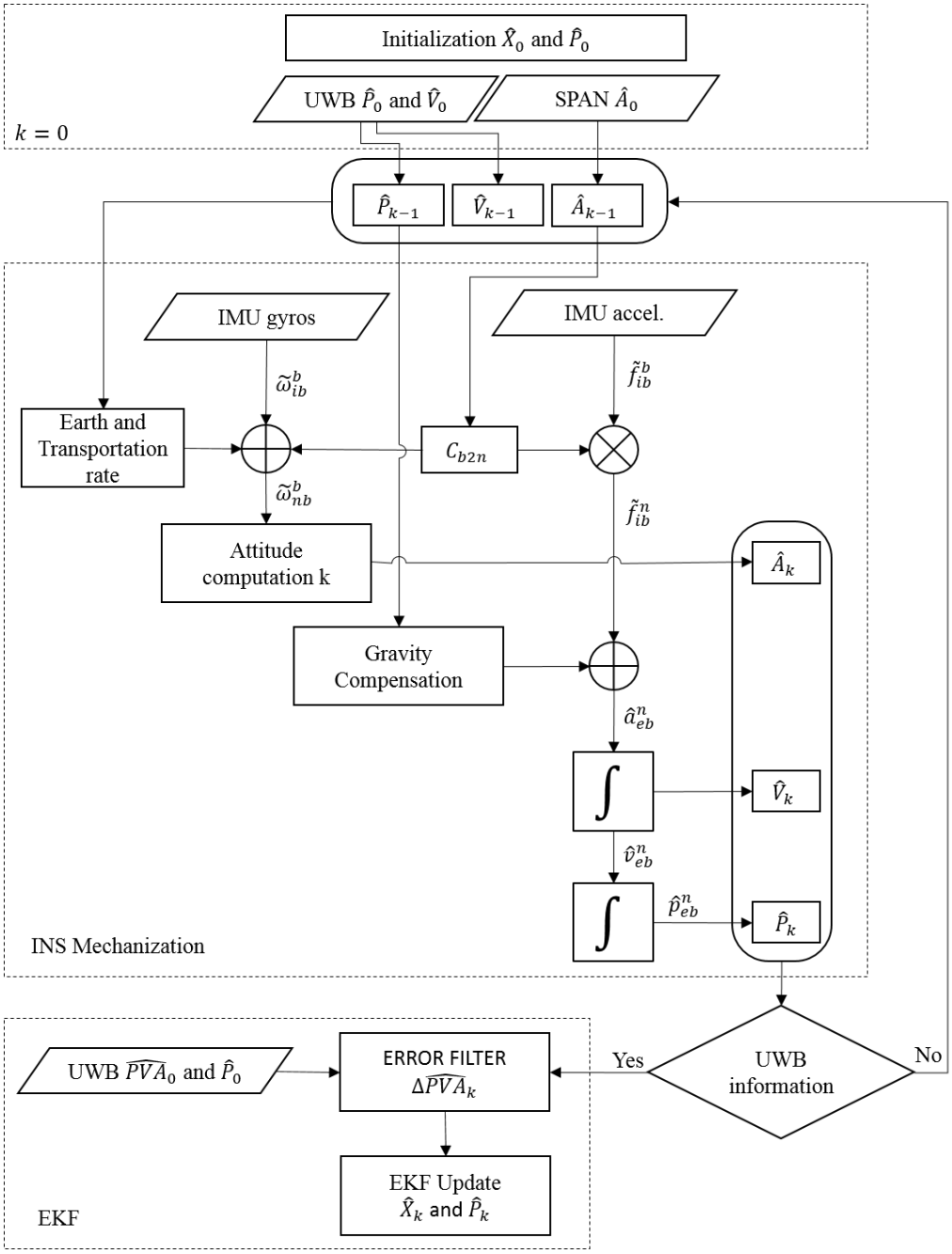


Figure 5.5: UWB/INS Data Fusion algorithm.

5.2.2.2 IRsensor-based visual odometry

However, the performances of these system suffers from noise and biases that results in large position errors increasing with time. Moreover, a data packet loss from the UWB network increase the errors in the estimation procedure. Such error can be mitigated using information about the system motion obtained from consecutive images acquired by a vision sensors. Unfortunately, as it has been said before, there usually is not enough light in mines for using normal camera so a very cheap infrared camera with 2 infrared LED attached to it has been used (Figure 5.3). The visual odometry (VO) algorithm used in this case is based on [8] where the classical scale problem is solved by pointing the camera sensor downwards at a fixed known height. The procedure start with the camera sensor acquiring a video during the motion of the pedestrian user. The camera looks at the ground during the motion acquiring frames at a fixed rate and features of the mine floor. These features are matched at each epoch with the correspondent features of the sequent image. With the assumption that all these features correspond to the reprojection of real 3D points belonging to the same planar surface (mine floor) is possible to affirm that an homography transformation relates the points in the two images. This homography can be computed using the DLT algorithm and has been already demonstrated his relation to the geometric parameters representing rotation and translation of the images. To estimate the correct homography transformation, the image correspondences must be cleaned by the presence of outliers i.e. wrong matches between features point. To do this, the DLT alogorithm is wrapped in a RANSAC framework and used to estimate the correct matrix. This matrix is used then as initial estimate to provide a staring point for a non-linear iterative minimization procedure based on the Sampson error. The homography matrix respect the relation in 3.2.1 and can be decomposed to find R and t between the images. Once that a new images is added the same procedure is repeated to estimate the new transformation. Then the procedure expressed in 3.2.1 is iteratively applied. All the processed is summarized as:

- Feature extraction in consecutive images with Speeded up robust features (SURF) approach [43].
- SURF descriptors, are match, namely identify the corresponding features in the consecutive images. i.e. the descriptors with minimum Euclidean distance.
- Mismatched features have to be discarded from the computation in order to avoid errors in the resulting motion and therefore Random Sample Consensus (RANSAC) [29] method deleting the match outliers is used during the homography matrix estimation.

- Finally, the resulting set of matched points are used for computing the motion from the homography rule with Levenberg-Marquardt's algorithm [36].

The visual odometry results are position and attitude of the camera center in each acquisition time (15 Hz). This information can be used then in a fusion algorithm that integrate the camera measurements with the INS/GNSS or INS/UWB implementation. The fusion algorithm is based on the Extended Kalman Filter implemented in a loose coupling approach, where estimates of the visual odometry phase are used for the measurements update. The state vector in this case will be composed by position, attitude, camera scale factor (solved knowing the fixed height of the camera) and IMU biases. The measurement vector instead is composed by VO position and attitude, and vision based velocity.

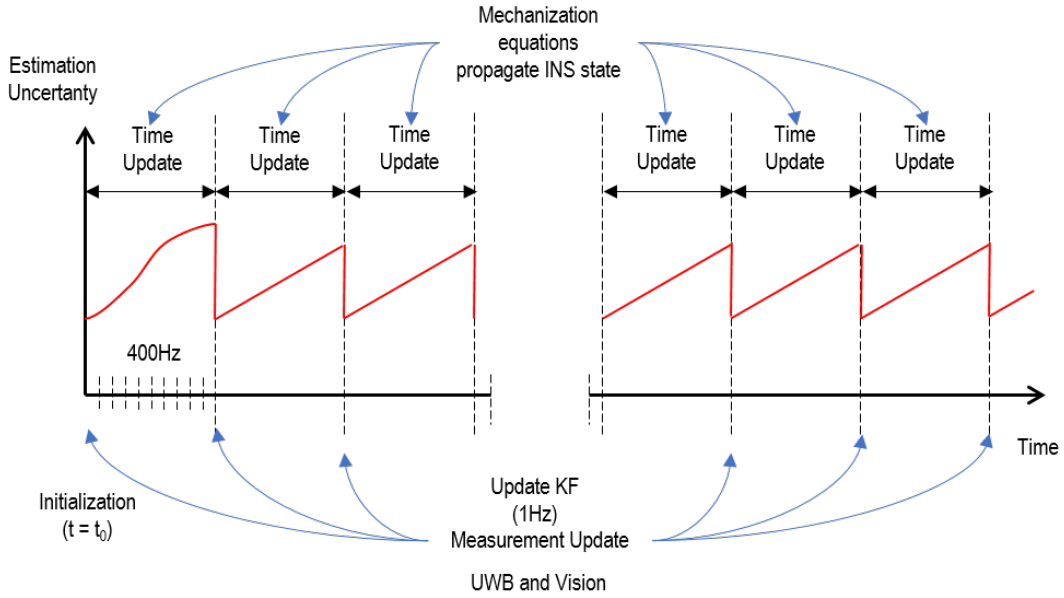


Figure 5.6: Extended Kalma Filter time update.

5.3 Case study: Otaniemi tunnel

The data acquisition has been performed in two different campaign during spring 2018 in Otaniemi, Helsinki (Finland). In the first test, the platform setup 1 has been used while in the second campaign the final platform setap has been tested (Figure 5.3). The test site was a tunnel of about 200 meters long and 5 meters wide. The path followed was a round trip path along a straight line. The pedestrian user has walked along the path holding the platform with the sensors in his hand. A backpack with a tactical grade inertial sensor on it has been put onto the user's shoulders. The data acquired by this sensor have been used as a reference solution.

In order to have a georeferenced reference solution, the sensor was initialized outside the tunnel and then brought to the test start point where the user was waiting. Before starting the acquisition, the network of UWB has been located in the tunnel and measured with a measurement tape. Images, inertial measurements, ranges and magnetometer information have been acquired during this path for post-processing (Figure 5.7).

Figure 5.7: Test performed in Otaniemi tunnel - Helsinki - Finland.

The test acquisition can be summarized in:

1. a network of UWB is located in an underground mine ensuring an accurate measurement of the antenna phase center position. This can be made with a integrated topographic survey made with geodetic GNSS receiver and a Total Station. In this way, the UWB network can be georeferenced in a geographic reference system. In the presented test, the topographic survey was avoided due to time constrains and the network was measured with a measurement tape;

2. when the UWB tag enter inside the network receive as first message an initial approximate position in the global reference frame (or local) and an initial approximate orientation defined a-priori;

3. the platform, composed by the previous described sensors, acquires range information, inertial measurements and infrared images during the pedestrian motion of the user;

4. all the measurement are integrated in post-processing combing the two algorithms core: the visual odometry and the navigation hybridization.

Particularly important for the set-up of the entire experimental test was the choice of the UWB TAG location along the tunnel. An optimal distribution should take into account both the ranging capability of the system and the geometry of the network. In the presented test, taking into account the coverage of the UWB network evaluated in previous tests, 8 anchors have been located along the tunnel (4 anchor for each side). Figure 5.8 shows the UWB network.

Figure 5.8: Otaniemi Tunnel. Helsinki, Finland.

What is evident from this network configuration is that the anisotropic geometry doesn't allow a good conditioning of the parameter estimation. This situation is reflected mathematically from the difficulty to invert the design matrix which can be proximal to the singularity. Unfortunately, this problem is not avoidable as it require the increase of the number of anchor located along the tunnel, anchor that wasn't available during the test. Therefore the UWB positioning performed with classical minimization problem could lead to bad estimates as will be shown in next section. An alternative to this problem is performing other kind of estimation procedure as will be shown later.

Table 5.3 summarize the characteristics of the two test performed during the data acquisition in Otaniemi tunnel.

Table 5.3: Tests parameters.

	Test n° 1	**Test n° 2**
UWB		
Acquisition time	534,739 sec = 08:91 min	537,80 sec = 08:96 min
N. of estimates	1912	1893
Positioning rate	3,57 Hz	3,52 Hz
N. of anchor	8	8
Algorithm	2D Kalman Filter	2D Kalman Filter
INS		
Acquisition time	548,98 sec = 09:15 min	578,13 sec = 09:63 min
Measurements rate	400 Hz	400 Hz
N. of estimates	219595	231252
Videos		
Acquisition time	437 sec = 07:17 min	426,13 sec = 07:06 min
Frame rate	15 Hz	15 Hz
N. of frames	6555	6391
Video resolution	1920 × 1080	1920 × 1080

5.4 Fusion algorithm validation and results

In this section, the details of the algorithm implementation and the relative problems will be discussed. Starting from the IR-based Visual Odometry algorithm, the issues and the solutions adopted will be presented. At the end, the validation of the algorithm and the results will be discussed in therm of navigation quality. Then the section will move on the discussion about the UWB/INS hybridization algorithm. The problems related to the data synchronization, measurements errors and integration will be deeply analysed. The obtained results will be presented and the navigation solution will be estimated. In conclusion, the fusion of this two main cores will be discussed and final results will be validated.

5.4.1 IR-based visual odometry results

This section discuss the the visual odometer algorithm and especially the challenges in estimate the translation between consecutive images in underground mines. As stated in Section 5.1.1 the main challenge in an underground environment consist in extract features from images acquired in low-light condition. As the RGB imaging sensors can't acquires visible light in this condition, the first proposed sensor to overcome this issue was the low-cost thermal imaging camera Flir ONE PRO. The idea was that, if the mine presents some temperature gradient, the thermal camera can register this variation as pixel radiometric variation (i.e. features).

Unfortunately this assumption was wrong. In fact the images acquired during the first test don't present any usable features. The reason is that the variation of temperature on the mine floor is mainly constant or under the sensibility level of the thermal sensor. Figure 5.9 shows the difference between a thermal image acquired outdoor vs an image acquired in the mine.

Figure 5.9: Comparison between thermal camera outdoor acquisition vs underground acquisition (Flir ONE Pro).

For this reason, the imaging sensor has been changed with a low-cost infrared camera (Night Vision "IR-CUT", Figure 5.3) already described in Section 5.2.1. The camera sensor is composed by two active IR led which illuminate the environment at a distance between 3m to 5m. The back-projected light illuminate the sensor and a gray-scale images can be acquired. Figure 5.10 shows the sensor and an image acquired in the mine. In this case is evident that a sufficient number of image features has been acquired. These informations are the primary data for the visual odometer procedure.

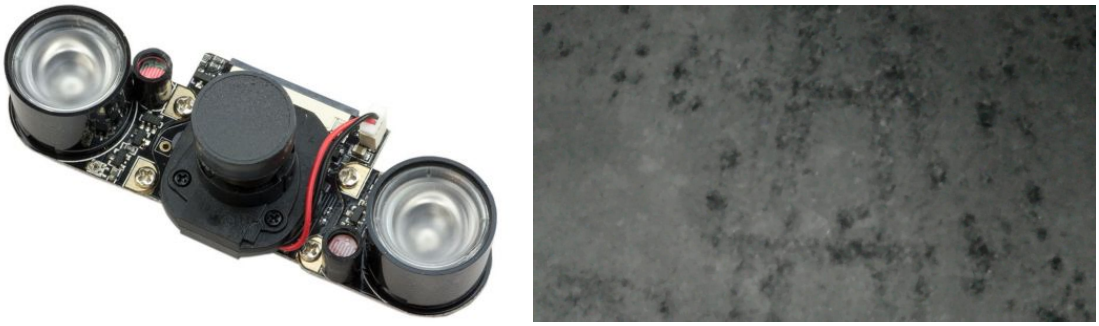


Figure 5.10: Raspberry Pi camera Night Vision "IR-CUT" and an example of image acquisition in mine.

To extract and describe features from infrared images, the Speeded up robust features (SURF) has been used. Unfortunately, as shown in Figure 5.11 after the motion start the images extracted by the video frames results heavily blurred. This is due to the fact that the sensor acquires frames while it is in motion. Ideally, if the video frame were acquired instantaneously allowing sufficient light to enter in the optic sensor, this problem could be avoided. In reality the characteristic of the sensor and the optics mounted doesn't allows this. Looking to them is possible to estimate roughly which is the expected motion blurring effect on the image pixels with the following relation:

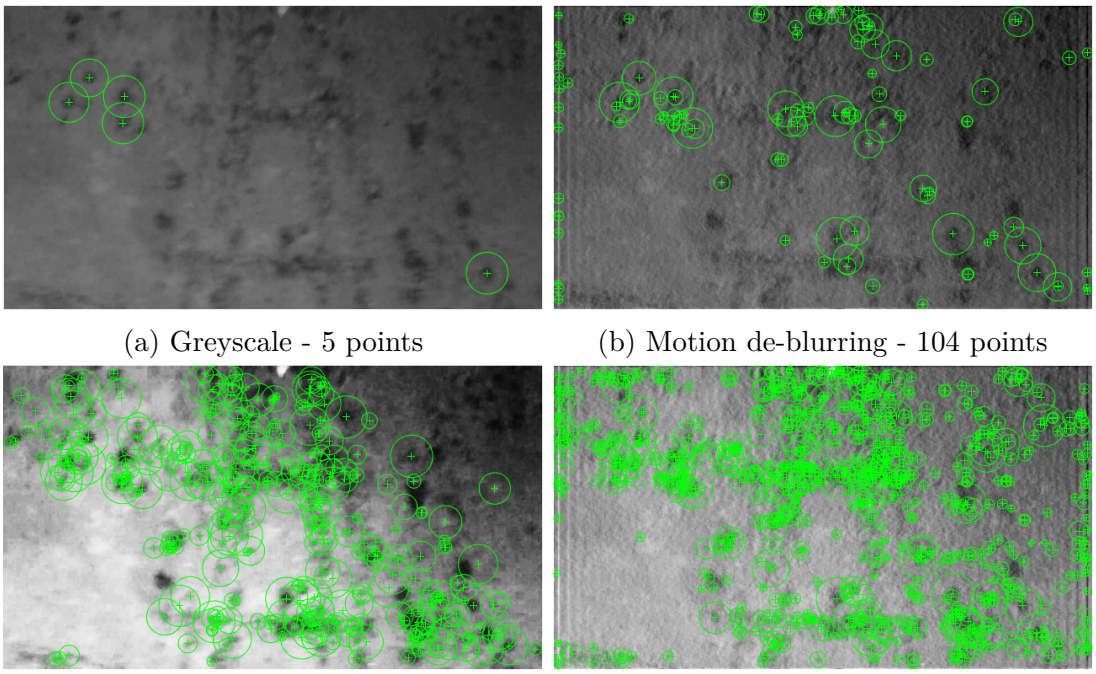
$$\lambda = \nu \cdot \tau \cdot \frac{c}{H} \tag{5.1}$$

Figure 5.11: Blurring effect during the camera motion.

where λ is the velocity of the camera in its translation motion, τ is the exposure time, c is the focal lenght and H is the distance between the image plane and the ground plane.

Assuming a walking velocity of 1 meter per second, with the camera sensor distant 1 meter away from the ground and given the characteristics of the sensor

shown in table 5.2, Equation 5.1 shows that on the floor, the expected image blurring effect correspond to 38 cm. In the image sensor it correspond to 94 pixel which is an extremely high value of pixel blurring. Having no possibility to reacquire the images changing some of the previous parameters, a pre-processing procedure for image enanchement and motion de-blurring has been investigated and evaluated. This procedure, based on the work of [104] allows to operate on the pixel value and obtain a better image. Figure 5.12 shows four different algorithm applied on the images and how they increase the capability of SURF to detect and extract a sufficient number of features. Applying a simple contrast enhancement algorithm the number of features increase from 5 to 354, while using also the motion de-blurring algorithm the number of detected features increase up to 909 (for the tested image).



(a) Greyscale - 5 points

(b) Motion de-blurring - 104 points

(c) Contrast enhancement - 354 points

(d) Both - 909 points

Figure 5.12: Image enanchement pre-processing. a) grayscale image, b) motion de-blurred image, c) contrast enanchement image, d) constrast enanchement plus motion de-blurring image.

This filter is applied iteratively at each epoch of visual odometry estimation in order to assure sufficient starting information for apply the procedure. The first epoch of acquisition requires an initialization procedure for further pose estimation. It consist in defining initial parameters (R_0, t_0, $C_0, attitude$) and identify the starting epoch of motion with an automatic procedure. For the first, the UWB initialization can be used as starting localization point. For the second, a median

of squared root distance between features points in two consecutive images is computed and compared to a starting motion threshold set to 5 pixels. Although this pre-processing procedure, the heavy blurring effect on the images acquiring during the motion doesn't allows to extract sufficient features to apply the proposed algorithm. This represent an important limit on the proposed procedure, a second data collection campaign is required to acquire better frames but unfortunately it was not possible.

5.4.2 UWB/INS hybridization results

In this section the UWB/INS Hybridization is analysed. As discussed in Chapter 3.4.4 the idea is to use UWB output instead of the GNSS PVT solution in a loosely coupling with the INS. For this reason the first step of this integration consist in analysing the PVT solution estimated by the network of UWB sensors. The Time Domain PulsON 440 UWB system is able to provide a real time estimation of the position using Nonlinear Least Square solver or a Kalman Filter-based solver. To this purpose the user should determinate the position of the network tag in X, Y, and Z dimensions. The parameters definition and the acquisition test has been made using the dedicated application programming interface called RangeNet. Then the onboard micropocessor installed in the UWB platform handles the communication, compute the position and send the results along the network. The two test made during the data collection campaign has been made setting the Kalman-filter solver in the UI.

Following section will shows the results obtained with this provided algorithm. As result of the estimation procedure, RangeNet software provides two main data files: a) the estimated position along the three local coordinates for each estimation epoch and b) a list of range measurements for each epoch of communication between the receiver TAG and the fixed anchors. The range measurements has been used to improve the navigation solution in an EKF procedure. This solution is later integrated in the UWB/INS hybridization (Figure 5.13).

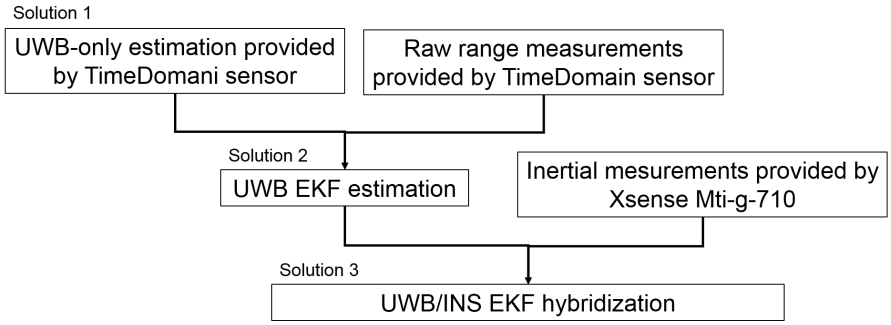


Figure 5.13: Workflow to obtain the UWB/INS navigation solution.

UWB-only Positioning Results from TimeDomain

Looking at the results, the first parameter to observe is the time of the estimation. Figure 5.14 reports on the X axis the number of estimation epochs of the UWB inner algorithm while on Y axis is reported the embedded clock time in first plot and the delta time in the second. From these plots is possible to observe some time gaps where the solution is not provided. In particular, during the Test n° 1, there are two peaks around estimation epoch number 800 and 1900 where the solution is provided after about 10 seconds. Moreover from the secon plot is possible to observe how the data rate is not constant with a mean value of about 3.57 Hz.

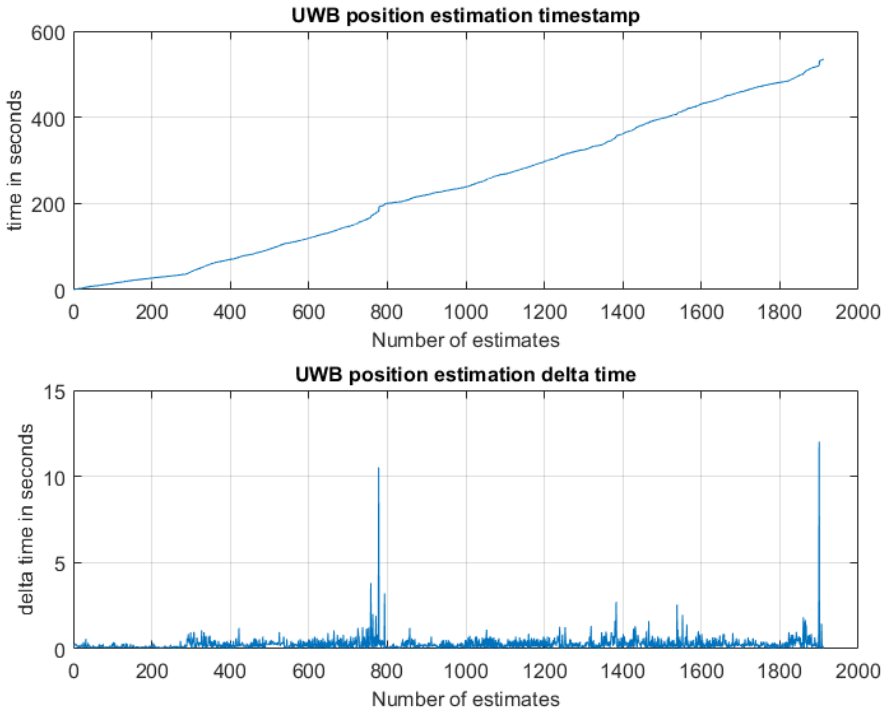


Figure 5.14: Time of the provided position and rate. Test n° 1.

Together with the results the software retrieve some flags to indicate if communications, ranging and positioning has been performed correctly or not, as well as Geometric Diluition Of Precision (GDOP) values and timing. Figure 5.15 shows four parameters to evaluate the quality of the result for the Test n° 1. Regarding the estimation method, the system indicates how far from the initialization the chosen solver has progressed. In this case, the Kalman Filter updating (blue diamond marker) has fail several times during the walking path. When this happens, the system re-initialize the estimates with a Non Linear Least Square (NLLS) approach (red diamond marker). Moreover, the system provide a Solver Error Code which indicates when the geometry of the anchor is to poor to support a NLLS solution (CODE 130 blue marker) and when the error estimation of the KF is too large

(CODE 132 red marker). The third plot of Figure 5.15 shows the GDOP while the forth plot shows the number of anchor used to estimate the position.

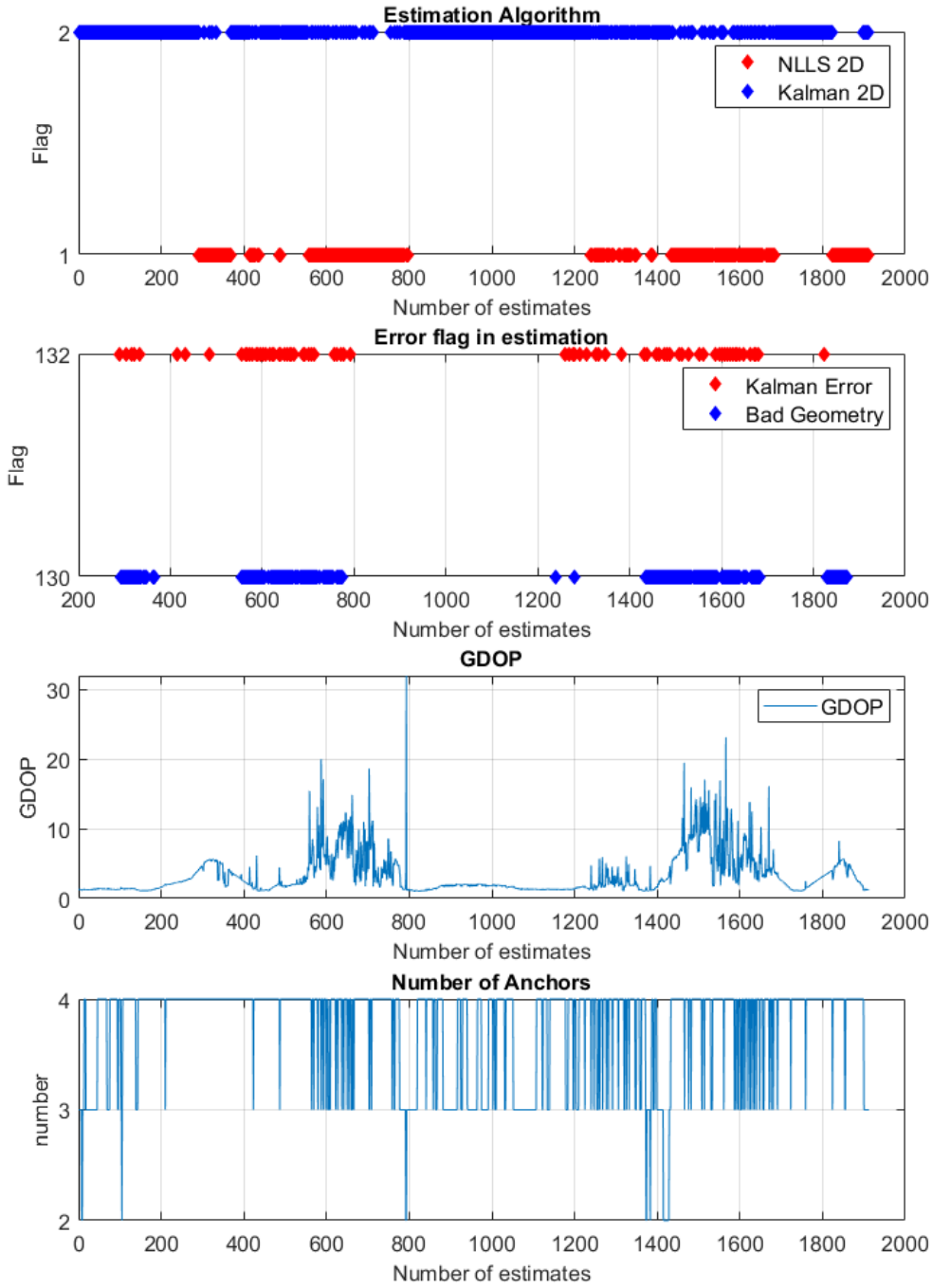


Figure 5.15: Information provided by the UWB CPU regarding the estimation procedure. Test n° 1

The results of the trajectory estimation provided by the inner algorithm are shown in Figure 5.16 together with the results cleaned from the erroneous solutions (Flags). From what has been shown previously is evident that the anisotropic geometry of the network is too bad to perform an acceptable navigation solution with the inner positioning algorithm.

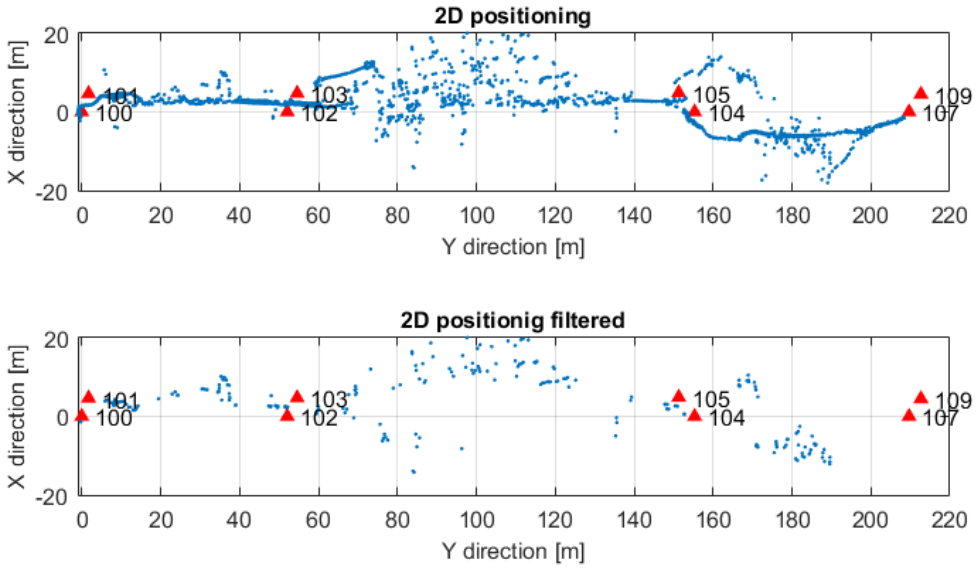


Figure 5.16: UWB inner solution.Test n° 1

The same consideration can been made for Test n° 2: Looking at the plots of this two test is possible to observe some common behaviours in the error flags and in the positioning. The errors are concentrated for both tests in the same portions of the path, upholding the "bad geometry" statement. The mean update rate in this test is 3.52 Hz with some data loss of about 10 seconds.

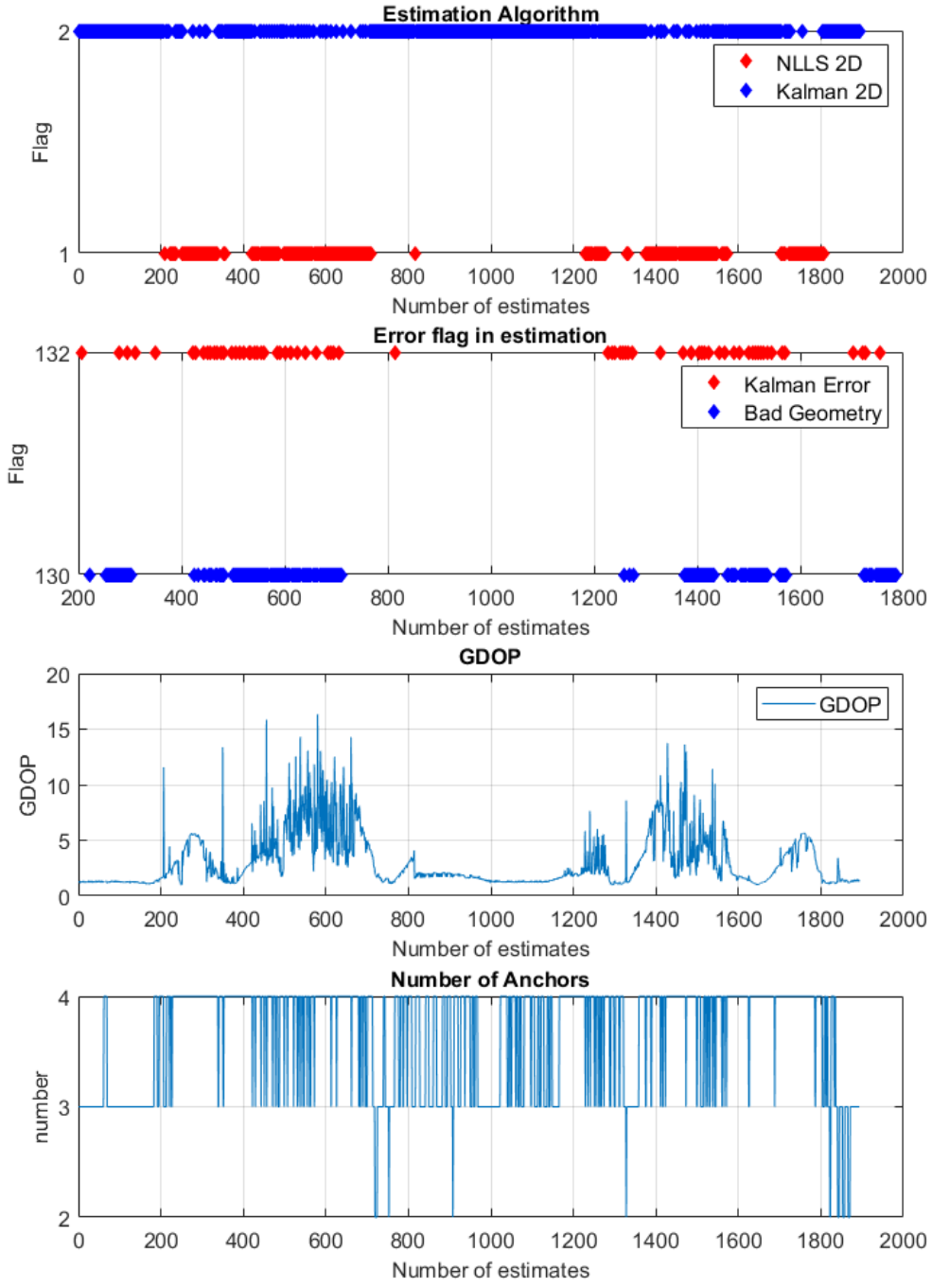


Figure 5.17: Information provided by the UWB CPU regarding the estimation procedure. Test n° 2.

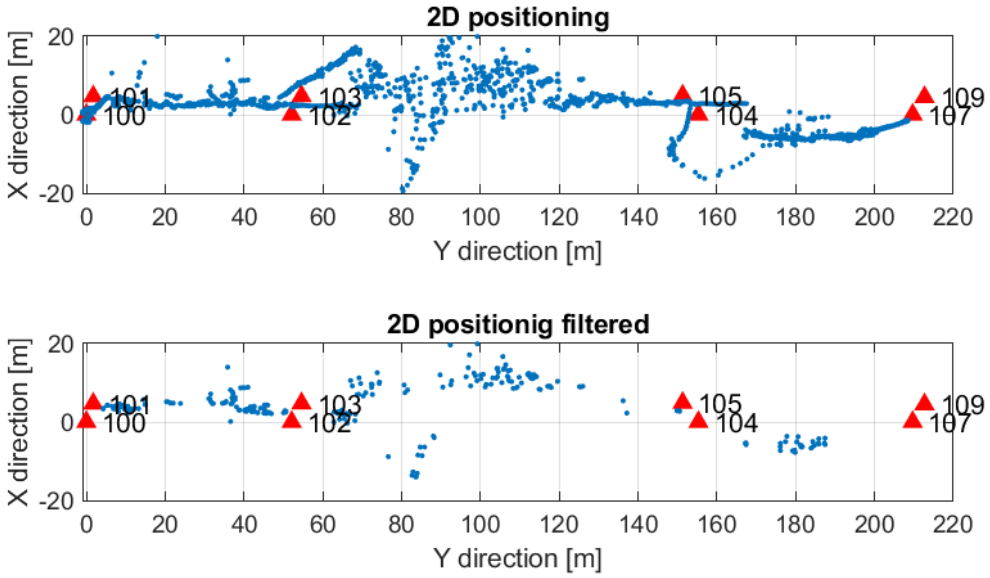


Figure 5.18: UWB inner solution. Test n° 2.

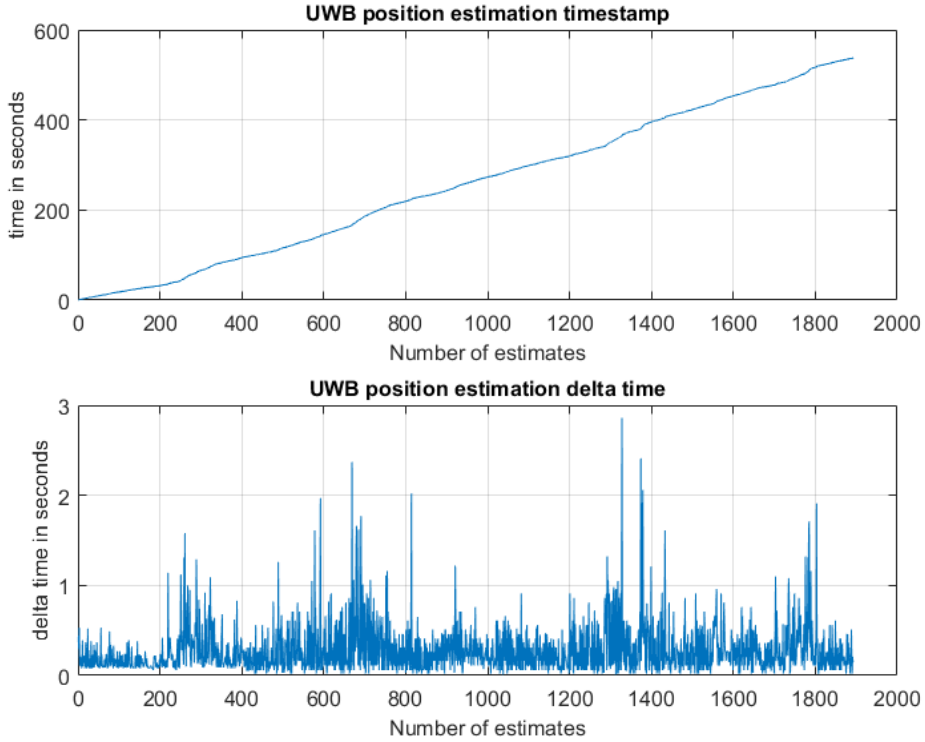


Figure 5.19: Time of the provided position and rate. Test n° 2.

Ranging Acquisition from TimeDomain

Together with the estimates, the sensor is able to acquire and provide range measurements used for the positioning. There are three different types of ranges:

- **Precision Range Measurements (PRM)** are taken using the TW-TOF ranging technique. These readings typically have high accuracy and are provided with estimates of range error as well as flags that warn of possible errors. The user can use these range error estimates to drive a Kalman Filter. The flags can be used to disregard inaccurate readings.
- **Coarse Range Estimates (CRE)** are analogous to RSSI (received signal strength indication) range estimates produced by continuous wave RF ranging systems in that they relate the strength of the received signal to range. They are different in two important ways. First, the signal strength reported is based on the strength of the first arriving energy and not on the strongest overall energy. This ensures that large signals produced by constructive multipath do not introduce false readings. Second, the signal strength reported is automatically calibrated based on the last successful Precision Range Measurement.
- **Echo Last Range (ELR)** measurements are Precision Range Measurements which have been taken between two other radios in the system. In other words, any time a unit initiates a range request, it will broadcast the last range measurement it successfully completed. For example, if Unit A measures the distance between Unit A and Unit B, it will broadcast this range measurement to Units C, D, E, etc., whenever it next initiates a range measurement. This is an alternate way of automatically distributing range information through a system.

Looking at PRM time acquisition, there are no data missing in the range measurements and the data rate is about 6.2 Hz for Test n° 1(Figure 5.20 and 6 Hz for Test n° 2 5.21).

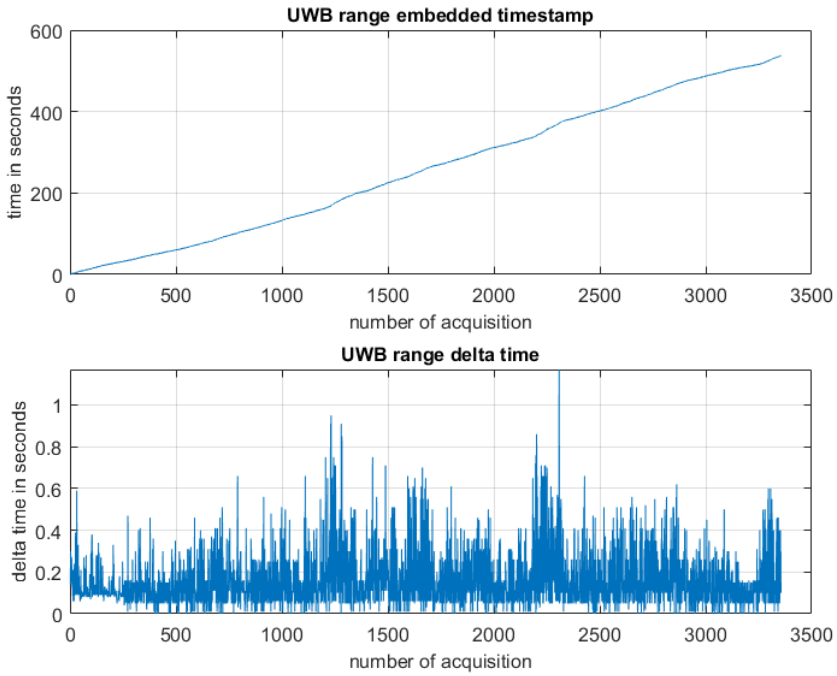


Figure 5.20: Range acquisition time. Test n° 1.

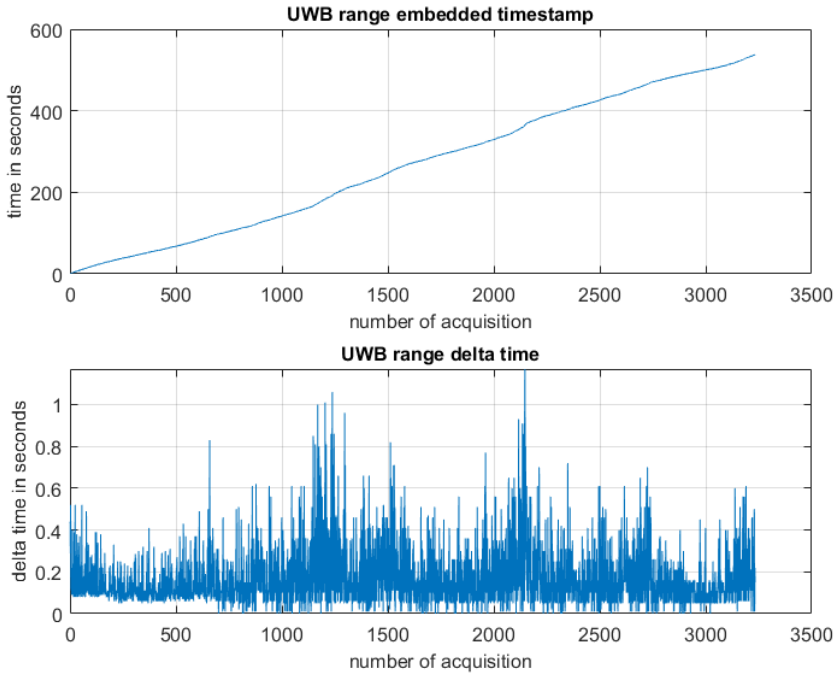


Figure 5.21: Range acquisition time. Test n° 2.

5.4.2.1 Implemented UWB EKF positioning

Acquiring raw range measurements allows to implement a more sophisticated algorithm with respect to the one proposed by the manufacturer. The raw ranges measurements has been used to perform an EKF independent from the inner algorithm proposed by the TimeDomain P440 system. Implementing an algorithm knowing the behaviour of the measurements and the geometry of the acquisition, could improve the positioning solution and increase the control on the estimation procedure. In order to do this, the procedure proposed in Figure 5.22 has been used. It consist in an initialization procedure which estimate the position of the UWB receiver during the standing part of the path. This initialization has been implemented in an automatic procedure which perform a non-linear least square minimization. Once the range measurements express the motion (raising a threshold), an EKF algorithm is used to estimate the trajectory. The update time has been setted at 1 Hz and the predictive model used to predict the state from previous estimate is supposed the uniform rectilinear motion.

Figure 5.23 is the ground trajectory estimation of the user walking along the tunnel that was equipped with the UWB sensor TAG. Figure 5.24 shows velocity estimation from which we can observe a walking velocity of about 1.5 m/s along Y axis, while about 0 m/s along X and Z axis. This is consistent with the ground track that shows a continuous motion along the Y axis. A great improvement of the positioning estimation with respect to the inner solution is evident from the results. It is possible also to note the absence of missing data value and outliers. Similar results are observable in the other test (Figure from 5.25 to 5.26). These outputs will be used as updating solution to integrate with the INS.

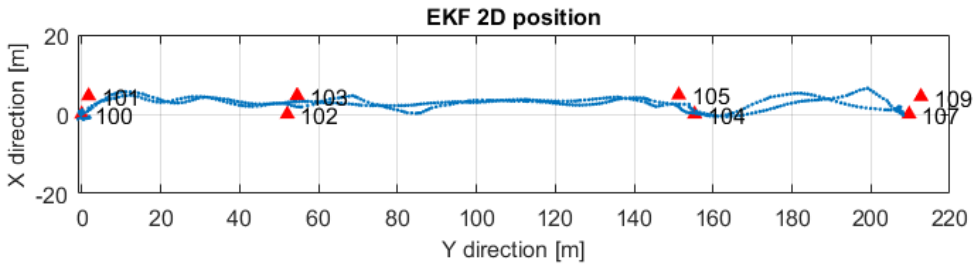


Figure 5.23: EKF navigation result. Test n° 1.

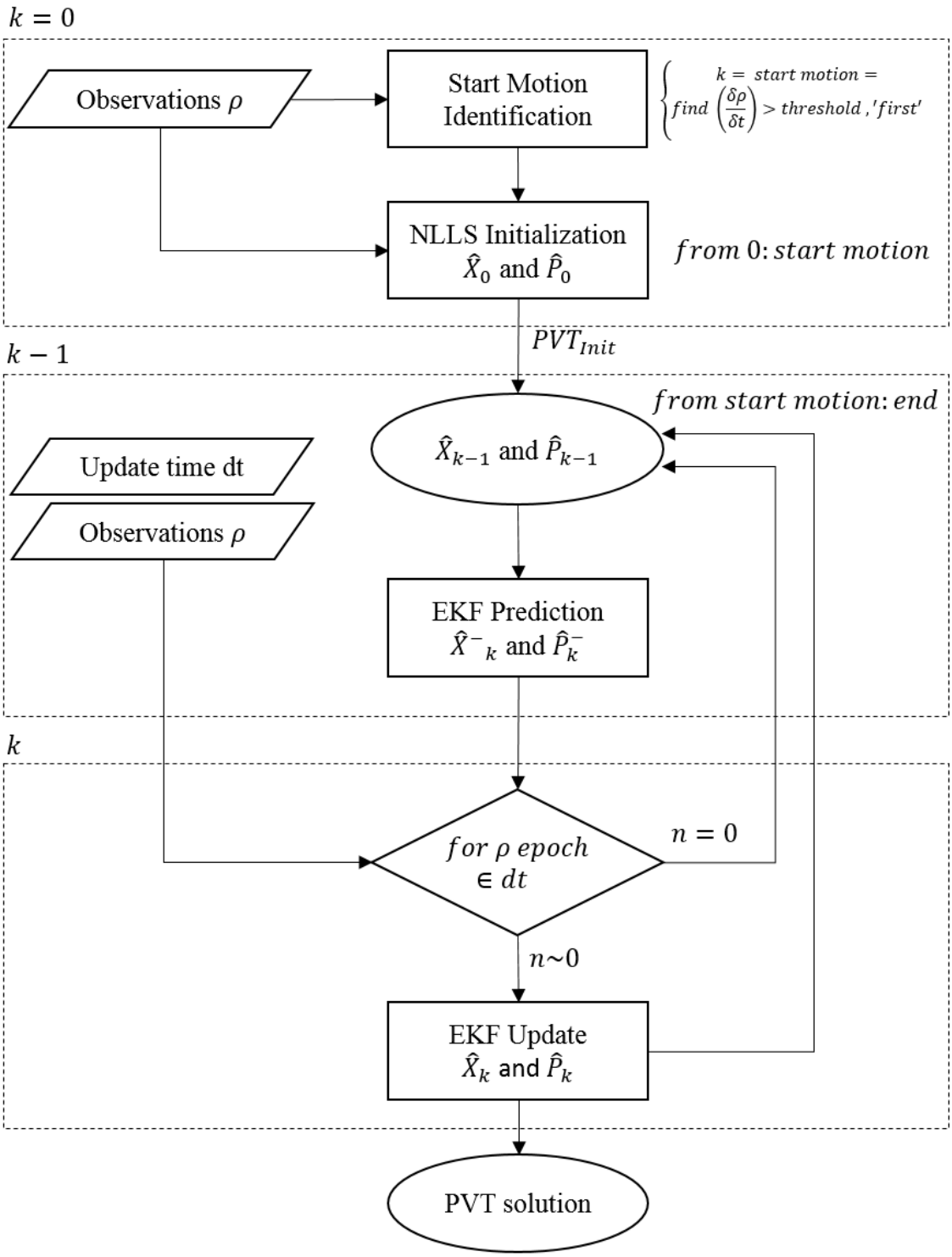


Figure 5.22: Flow diagram of EKF method used for UWB positioning estimation.

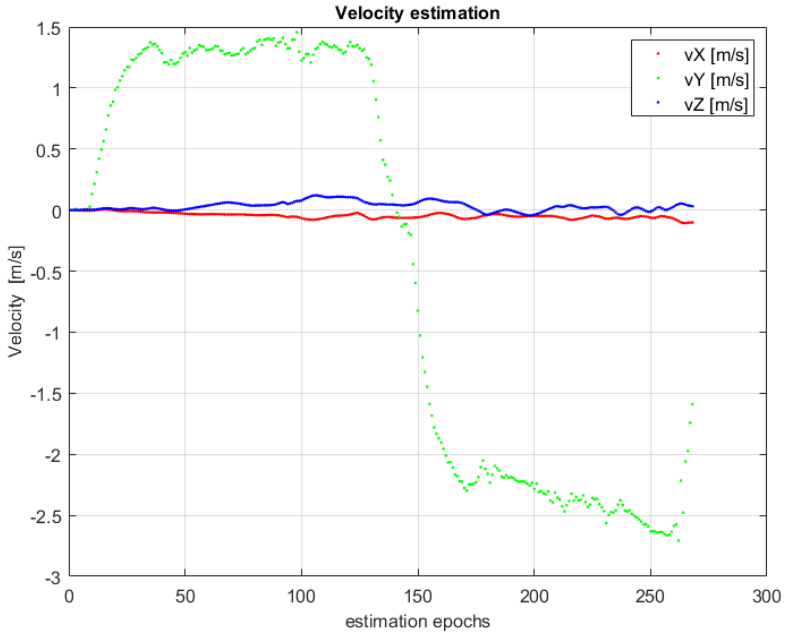


Figure 5.24: EKF velocity estimate. Test n° 1.

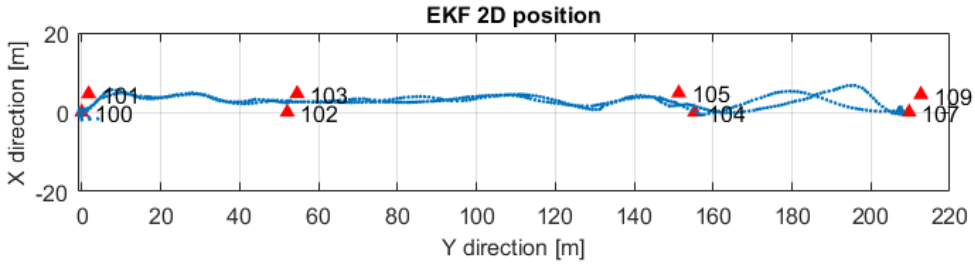


Figure 5.25: EKF navigation result. Test n°2.

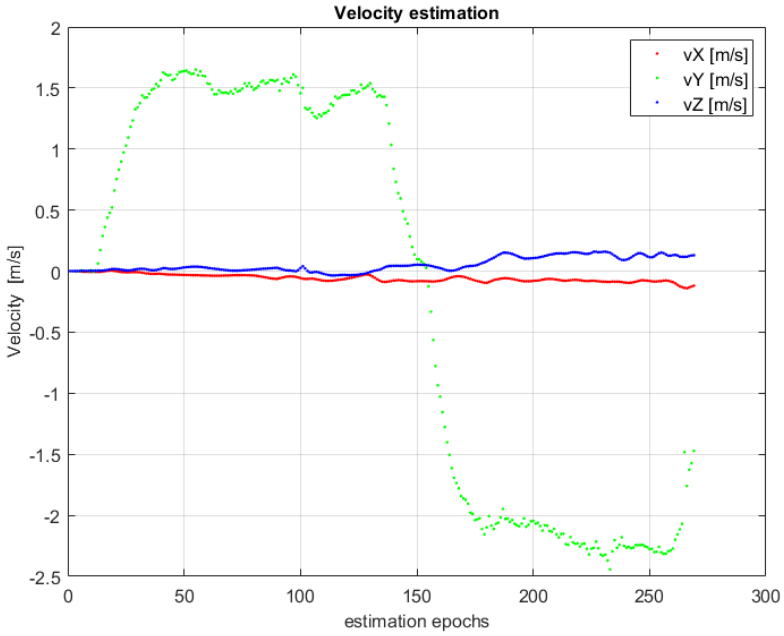


Figure 5.26: EKF velocity estimation. Test n° 2.

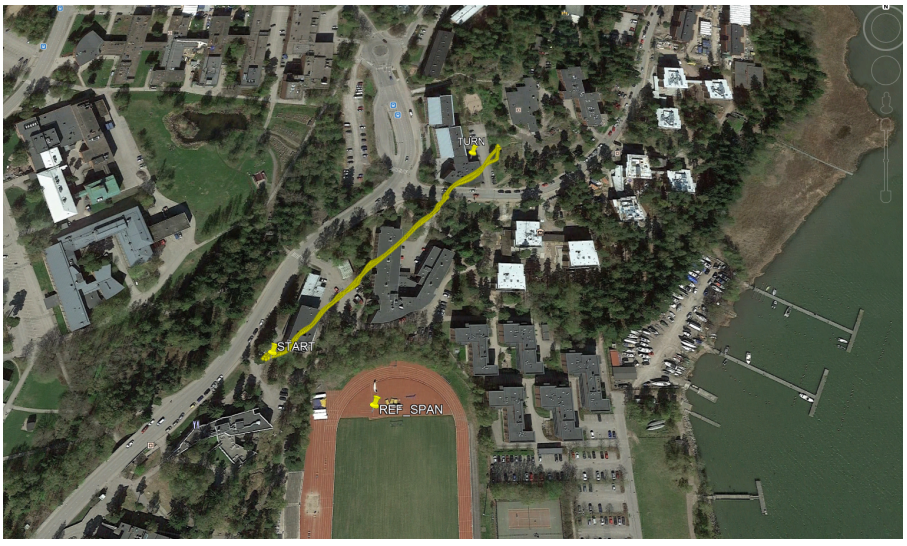


Figure 5.27: Position estimated with EKF for Test n° 1.

5.4.2.2 Inertial measurements analysis

In this study, acceleration and angular velocity measured by the IMU mounted on the platform are integrated sequentially to provide relative position and relative attitude information while the UWB position are used as absolute inputs for the

hybridization. For this purpose is important to characterize the noise level of the acquired raw data. With low-cost sensor is fundamental to perform static acquisition in order to compute measurement residuals [105]. As in the performed test an industrial grade INS is used, just the behaviour of the measurements has been analysed. In particular, looking at the following figure, is possible to observe two peaks in the rotation rate along the X axis, corresponding to a rotation of the platform performed by the user as a visual time stamp reference point. Looking at the gyros measurements along the Z axis is possible to identify a peak in the middle of the test. This is the turning point, when the platform, and so the user, has turned along himself to walk back along the path. The grey area correspond to some second of static position in the turn point before to move back along the path. The same information is provided by the relative Heading information, where at 360 seconds from the start, a turn of 180° is visible. Figure from 5.28 to 5.33 shows the measurements acquired for the two test performed. With this information and the UWB EKF estimate position and velocity, is possible to perform the INS/UWB hybridization.

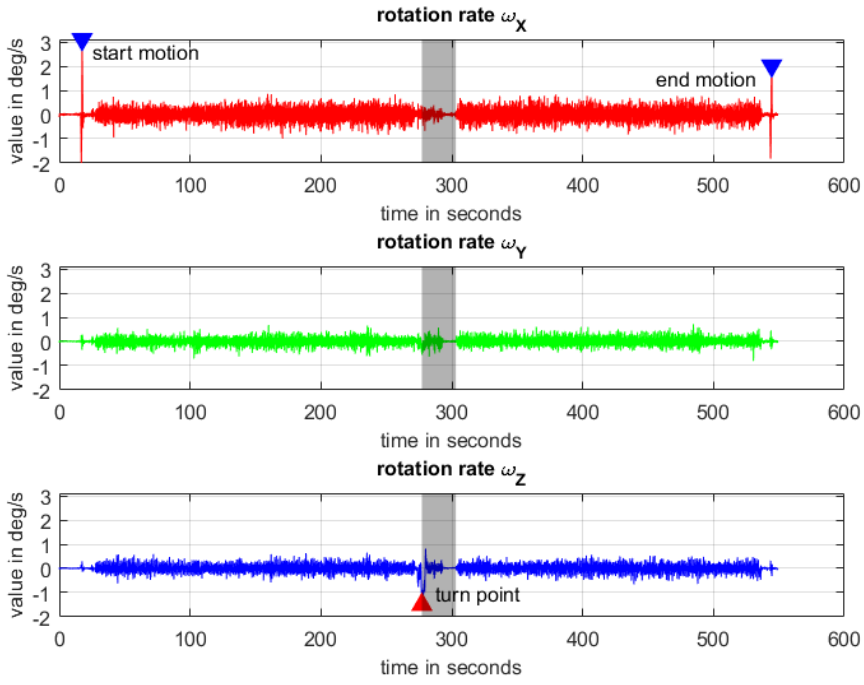


Figure 5.28: XSense gyroscopes measurements for Test N° 1.

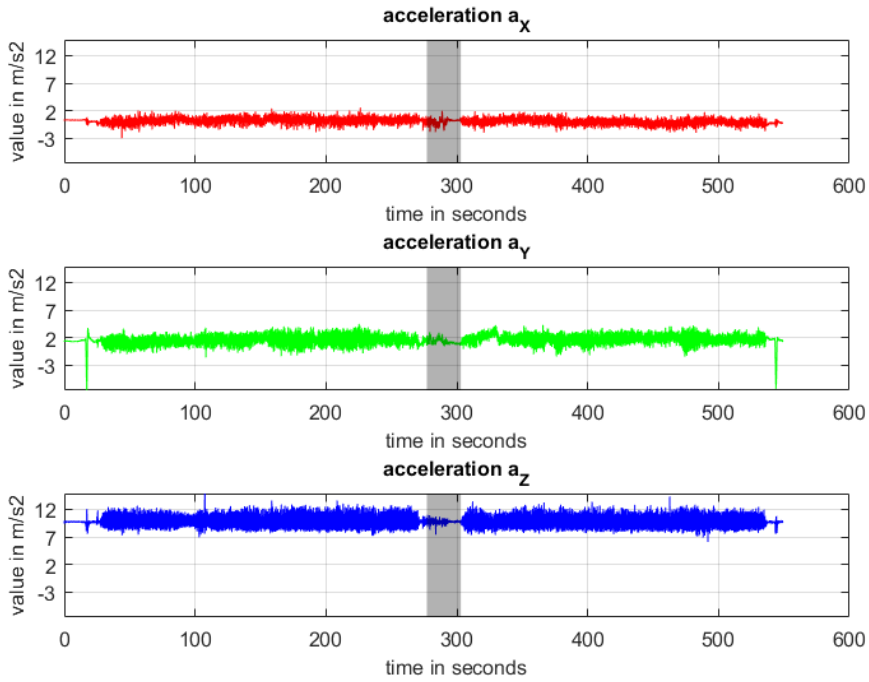


Figure 5.29: XSense acceleration for Test n° 1.

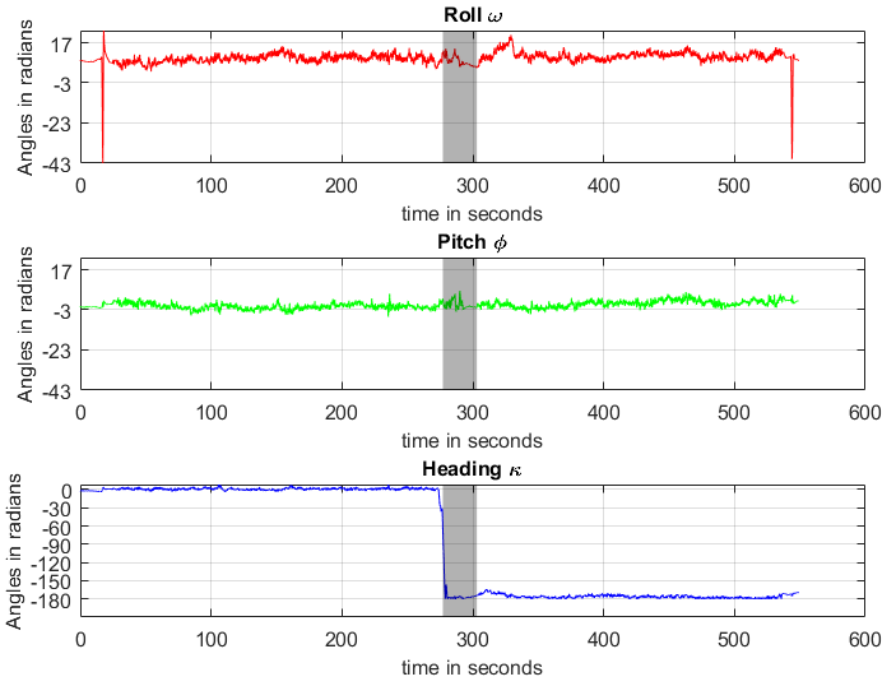


Figure 5.30: XSense attitude estimation for Test N° 1.

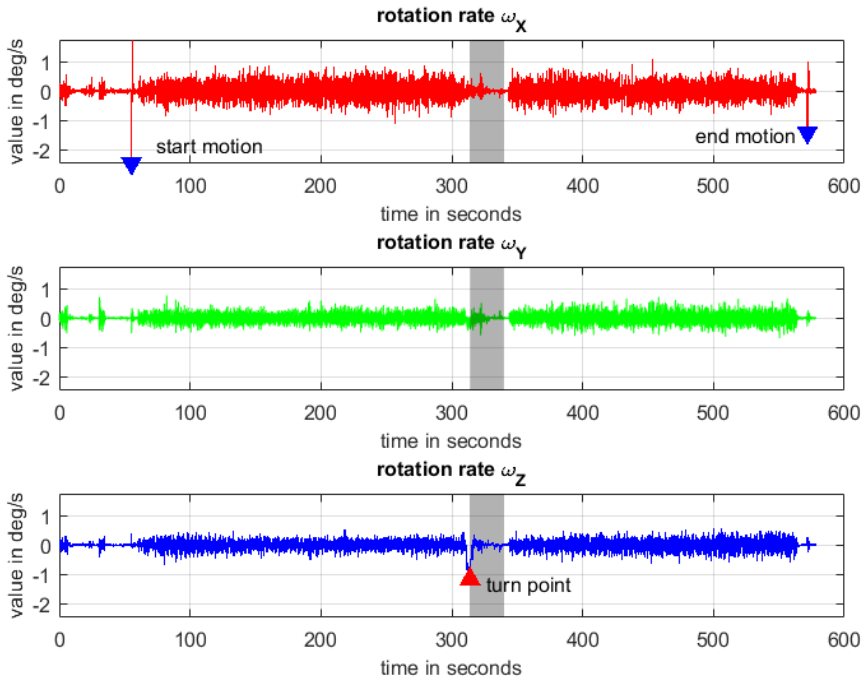


Figure 5.31: XSense INS data output. Gyroscopes measurements for Test N° 2. Identification of the turning point on the GyroZ axis.

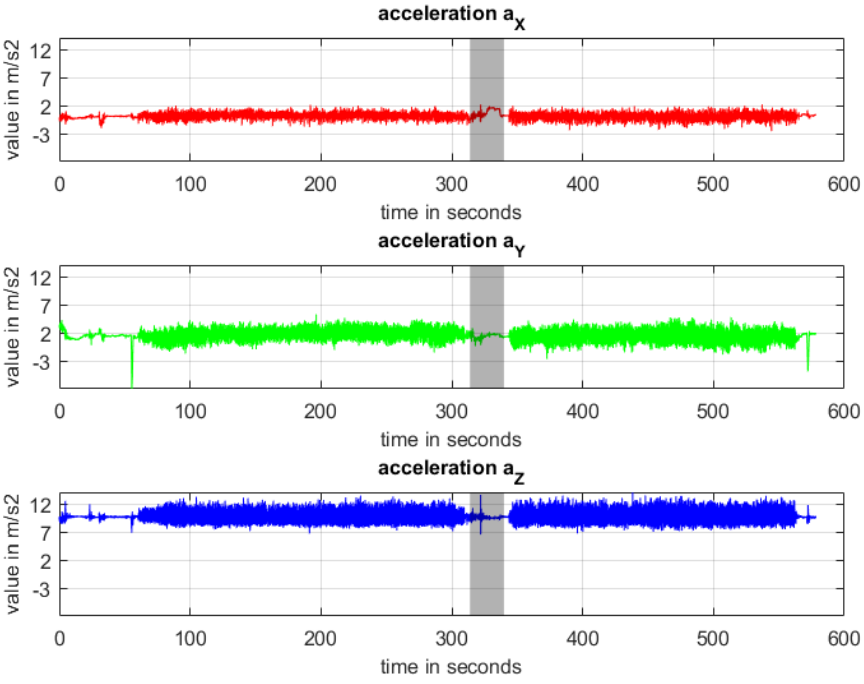


Figure 5.32: XSense INS data output. Accelereation along three axis for Test n° 2.

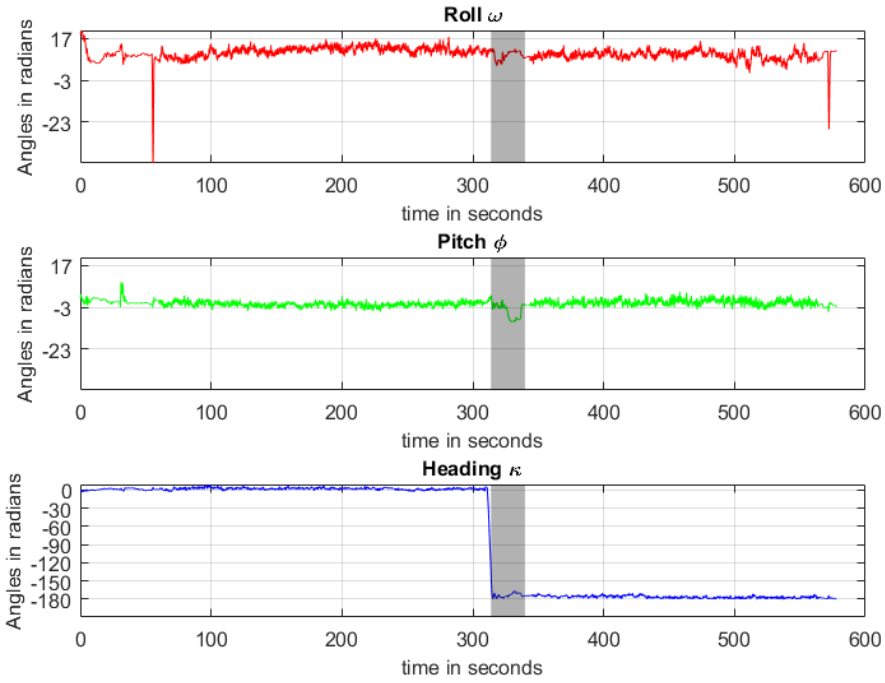


Figure 5.33: XSense INS data output Test N° 2. Attitude estimation.

5.4.2.3 UWB/INS Results

The theoretical concepts for the UWB/INS hybridization has been already explained in Section 5.2.2 and represented in Figure 5.4. Figure 5.5 goes more in deep in the description of the implemented algorithm. From the operative point of view, numerous problems have been addressed, mainly due to the time synchronization and the georeferencing of UWBmeasurements.

1. **Time synchronization.** Usually in a loosely coupled system integration, the problem of the unsynchronized clocks is addressed stamping the data received time of each sensor referred to a common clock and estimating the relative skew and drift rate. In this work unfortunately, each sensor system has acquired his data independently and with his own time scale. For the INS, the time is referred to the internal clock which start with the initialization of the sensor. The UWB measurements and the video are referred to the computational unit clock (PC and Raspberry respectively). It is evident, that without a common reference time, the time-related errors can decrease the accuracy of the navigation solution. To face with this problem, during the acquisition campaign, a rapid movement of the platform has been made by the user in order to generate a visual timestamp on the inertial rotation rate along the pitch direction. In order to do this, the differences between adjacent elements of the rotation rate has been computed and the first and

last peak of the resulting plot has been identified as the start and the end of the test. Another important information is the start of the movement which take place some seconds after the visual timestamp. In order to do this, a step detection algorithm has been used. It is based on the computation of the magnitude of the acceleration vector along the three axis. It show the general changes in acceleration. The plot shows that the acceleration magnitude is not zero-mean. Subtracting the mean from the data will remove any constant effects, such as gravity. Each peak on this plot (Figure 5.34) correspond to a step being taken while walking. This movement is easily recognizable also in the video frames. Using the visual marker (peak on the plot) the common staring movement timestamp has been identified for all the data. To validate the obtained result, this time identification has been compared with the GPS time provided by the SPAN system which has been initialized outside the tunnel. Finally, also the turning point of the walking path must be identified. This is easily detectable from the plot of the INS heading estimation as already shown in Figures 5.30 and 5.33.

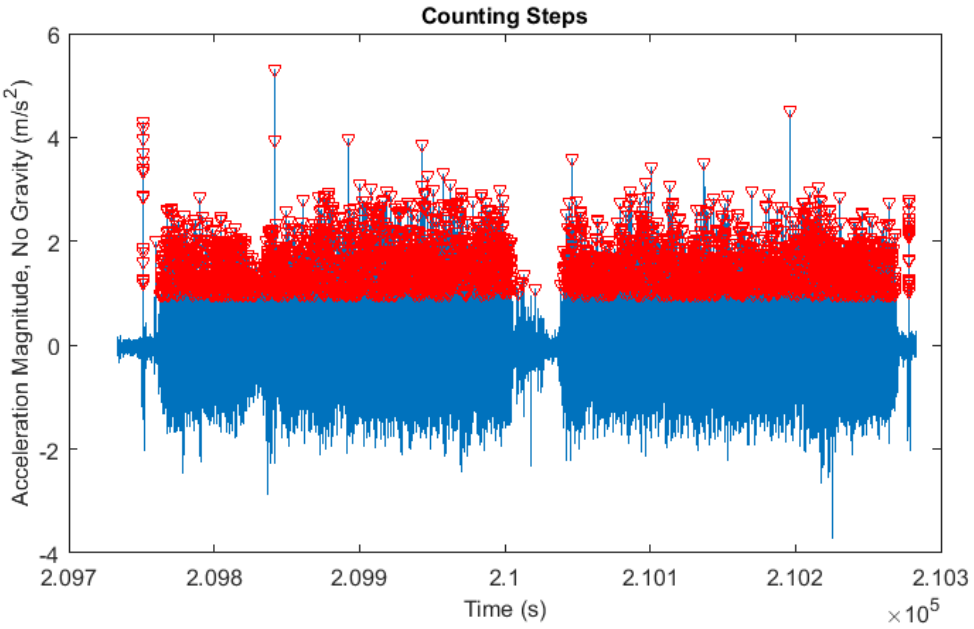


Figure 5.34: Step identification for start movement time.

2. **UWB EKF georeferencing.** In the loosely coupling of INS and GNSS measurements the mechanization equations of the inertial navigation are parametrized in north-east-down (NED) navigation frame. The entries of the position vector are the users geodetic coordinate in latitude (L), longitude (λ) and Altitude (h) which are converted thanks to the rotation matrix. Therefore in the proposed integration, the innovation process of the EKF need

to be estimated using UWB state vector in the correct georeferenced frame. Unfortunately, as no GNSS-topographic survey has been made during the implementation of the UWB network in the tunnel, the trajectory estimated by the UWB system is in local reference frame. To overcome this problem, the UWB estimated points have been converted from local to global reference system using an Helmert transformation and some known common points of the SPAN reference system [106].

These problems can introduce errors in the estimation procedure and should be avoided in future data acquisition campaign. Once addressed all these issue, the implemented algorithm has been used to process the data of the two tests and the results has been analysed. Figure 5.35 shows the results in cartographic reference system (ETRS89 - ETRS-TM35FIN) for Test 1. The red line represent the SPAN reference trajectory, the green points the estimate 2D position at the update time of the EKF. Figure 5.36 shows the same results in local NED system. This plot allows a metric visualization of the trajectories. Figures 5.37 shows the velocity estimates. Then Figures 5.38 shows the estimates of the accelerometer bias and Figure 5.39 represent rate gyro bias for the same time period (516 seconds of KF update). Finally Figure 5.40 shows the attitude history. The same analysis has been made for Test n° 2.

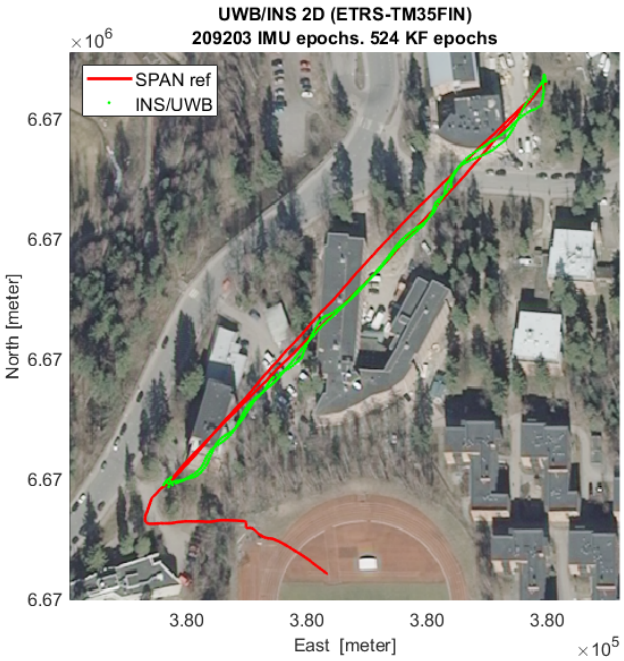


Figure 5.35: Navigation estimation in cartographic coordinates and reference SPAN solution. Test n° 1.

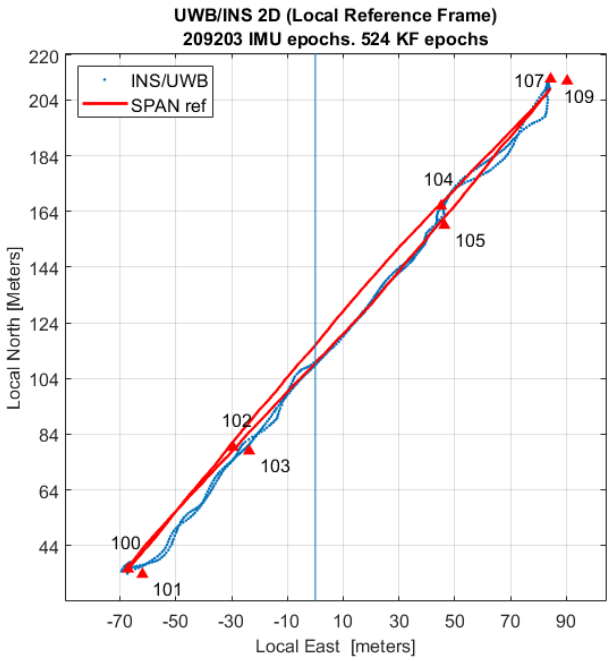


Figure 5.36: Navigation estimation in local NED coordinates (navigation frame) and reference SPAN solution. Test n° 1.

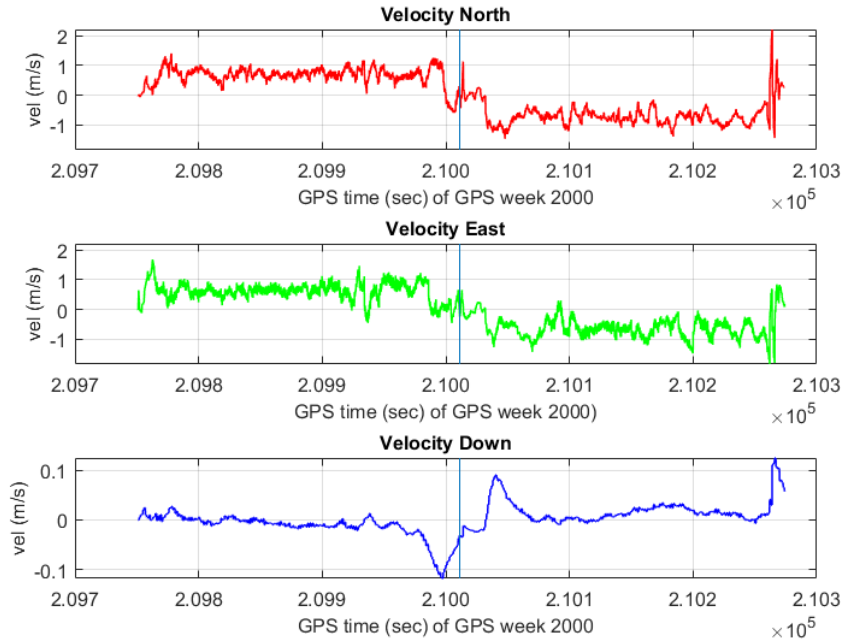


Figure 5.37: Velocity estimation from EKF along NED axis. Test n° 1.

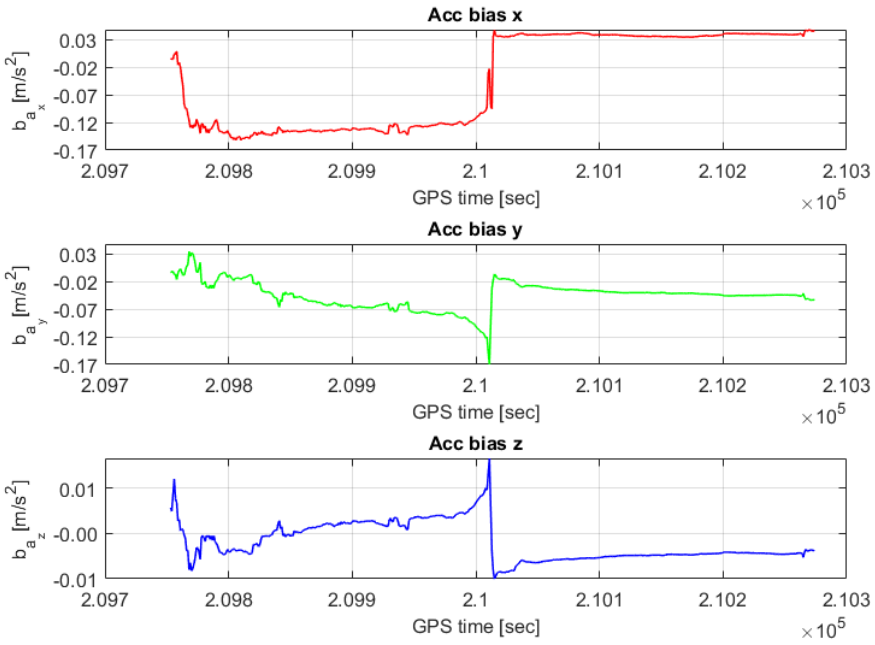


Figure 5.38: Sensor acceleration bias estimate for the low-cost UWB/INS integration. Test n° 1.

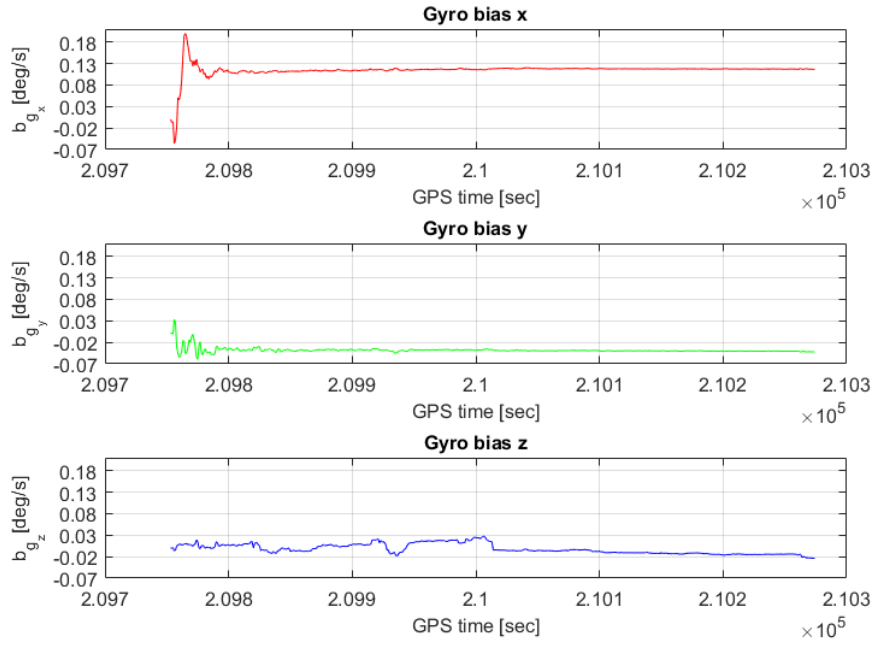


Figure 5.39: Sensor gyro bias estimate for the low-cost UWB/INS integration. Test n° 1.

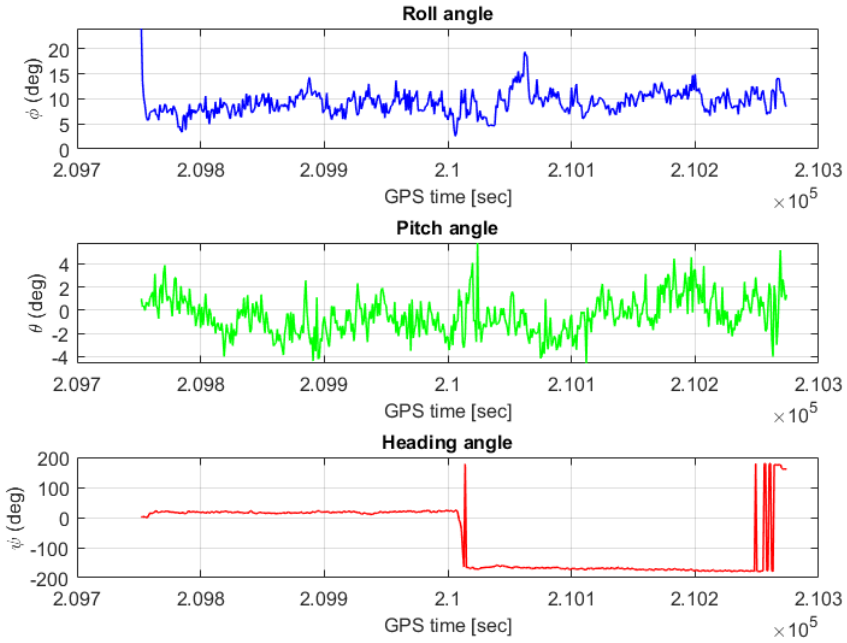


Figure 5.40: Attitude history for a low-cost UWB/INS integration. Test n° 1.

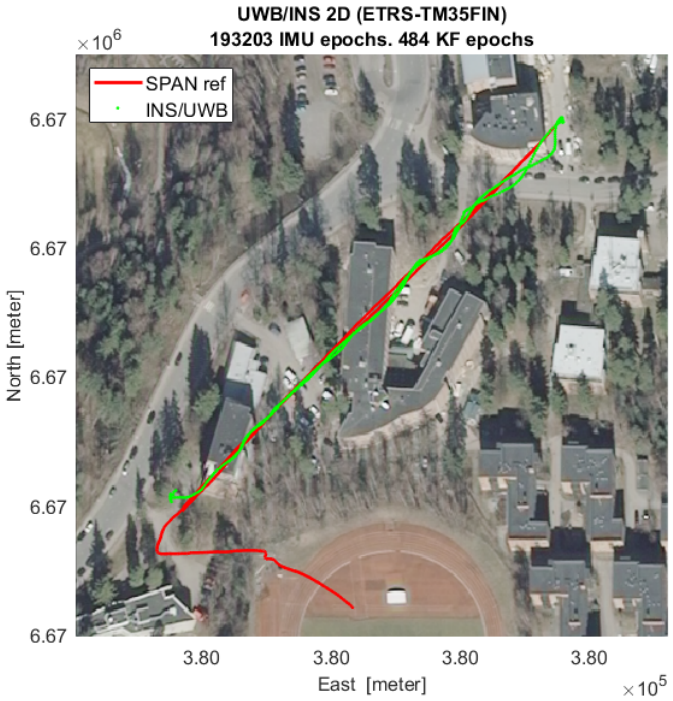


Figure 5.41: Navigation estimation in cartographic coordinates and reference SPAN solution. Test n° 2.

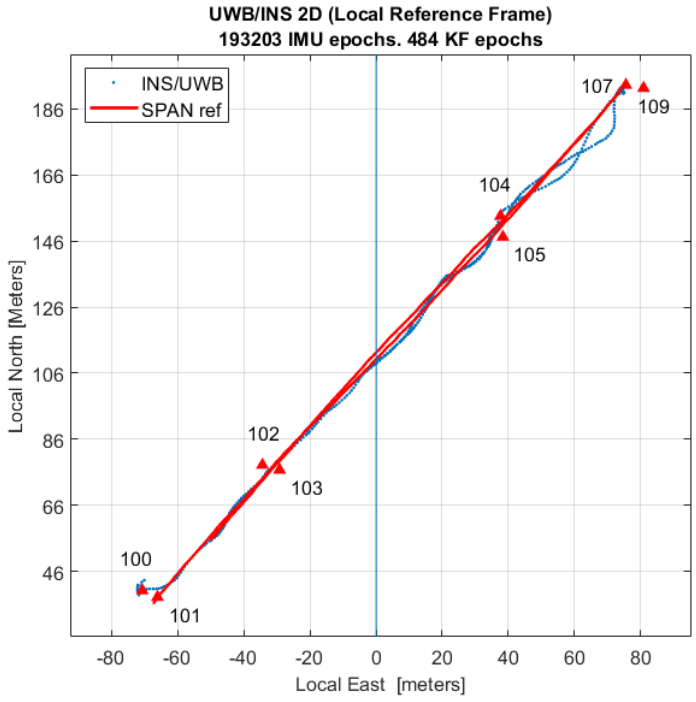


Figure 5.42: Navigation estimation in local NED coordinates (navigation frame) and reference SPAN solution. Test n° 2.

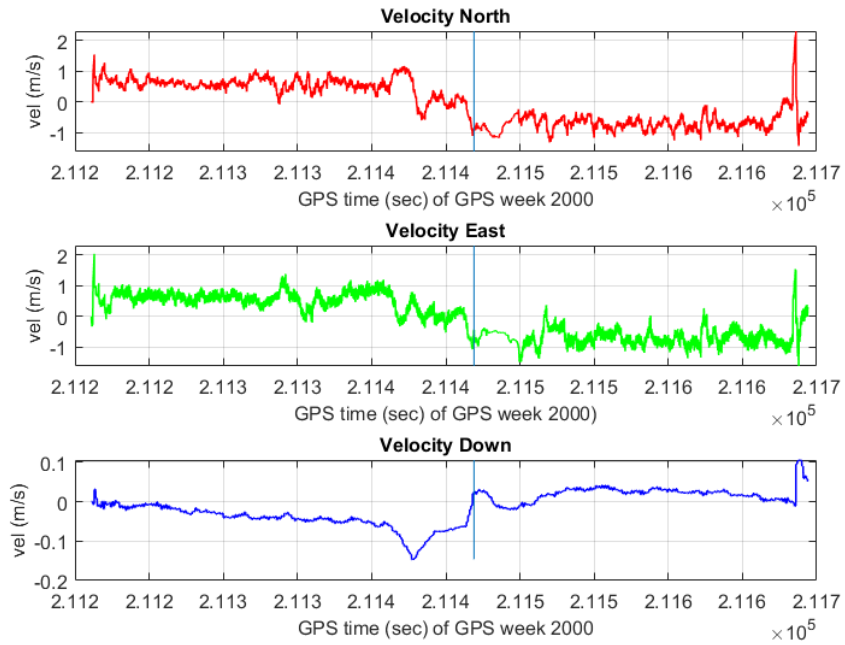


Figure 5.43: Velocity estimation from EKF along NED axis. Test n° 2.

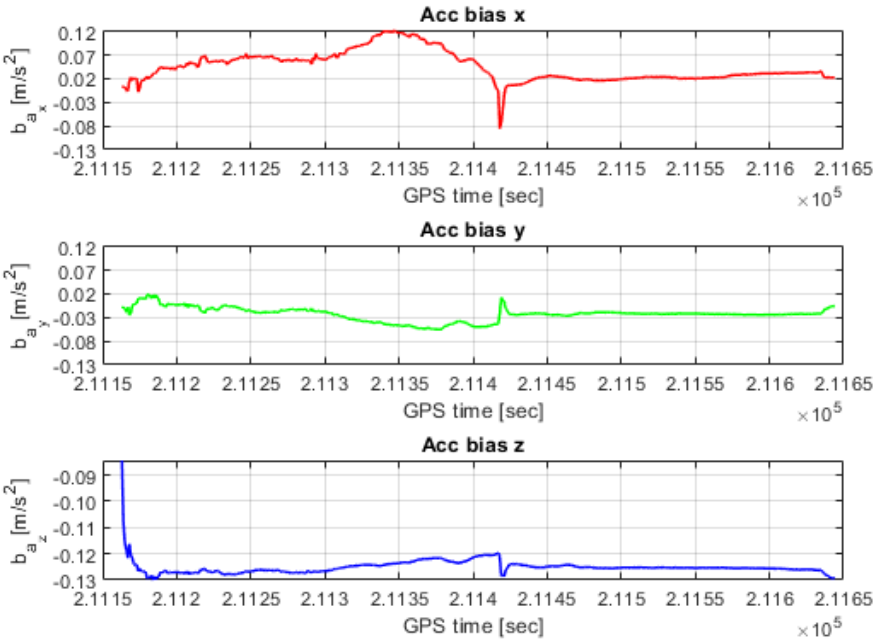


Figure 5.44: Sensor acceleration bias estimate for the low-cost UWB/INS integration. Test n° 2.

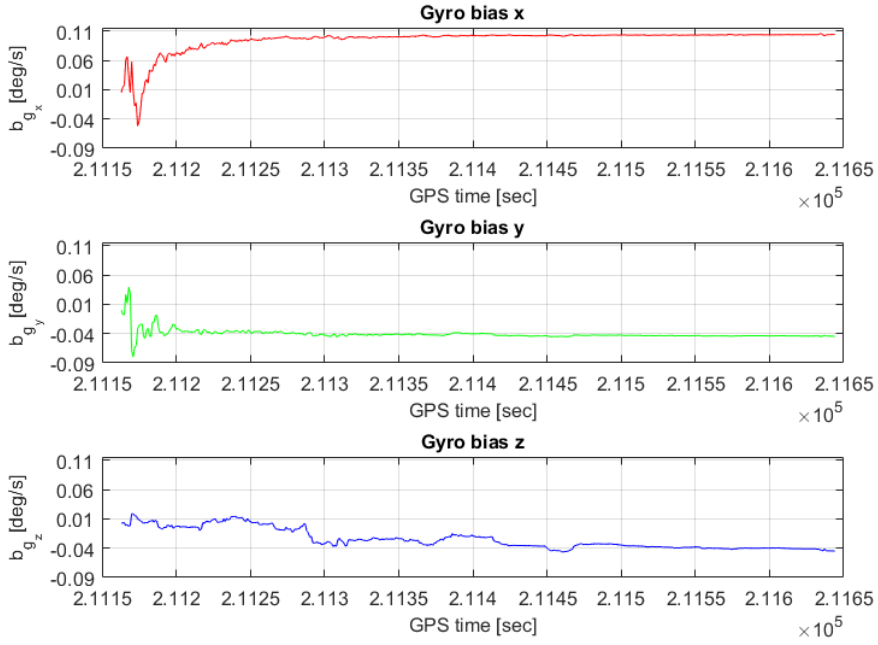


Figure 5.45: Sensor gyro bias estimate for the low-cost UWB/INS integration. Test n° 2.

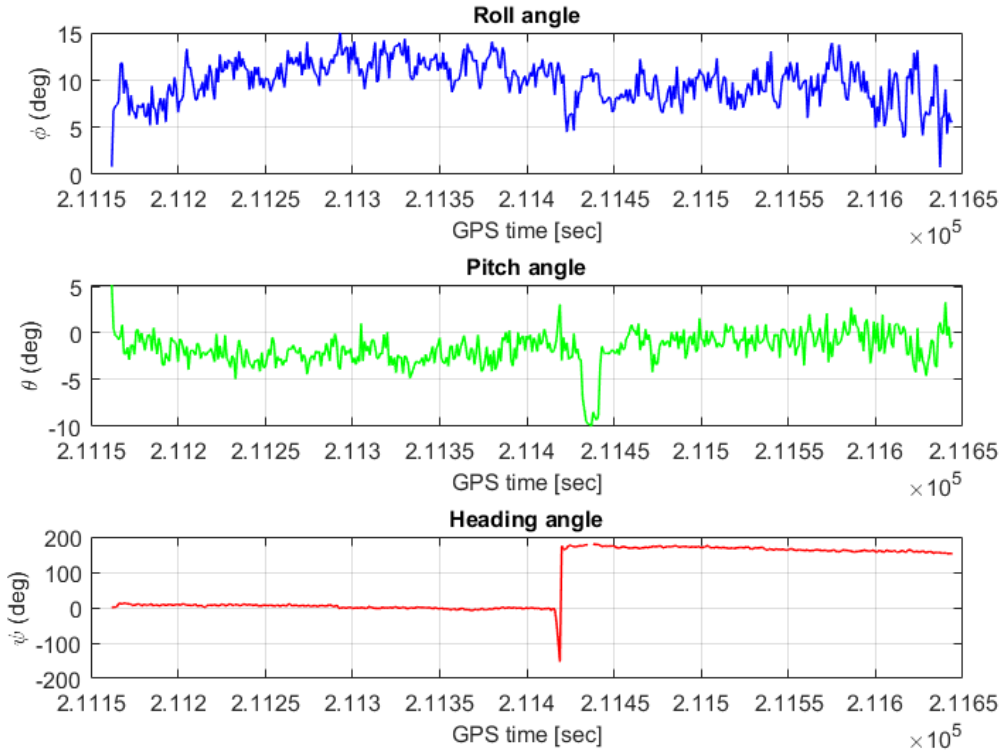


Figure 5.46: Attitude history for a low-cost UWB/INS integration. Test n° 2.

Finally, in order to validate the accuracy of the estimation, the positioning solution obtained was compared to a reference trajectory acquired with the SPAN system. These comparison has been made following the test procedures reported in [107] which estimates 3D and 2D error between the user position and the reference path at a given estimation time. The errors are defined as follow:

$$PE_{3D} = \sqrt{(\widehat{X}_{East} - X_{East})^2 + (\widehat{Y}_{North} - Y_{North})^2 + (\widehat{Z}_{Down} - Z_{Down})^2} \tag{5.2}$$

$$PE_{2D} = \sqrt{(\widehat{X}_{East} - X_{East})^2 + (\widehat{Y}_{North} - Y_{North})^2} \tag{5.3}$$

where $\widehat{X}_{East}, \widehat{Y}_{North}, \widehat{Z}_{Down}$ are the estimated coordinates and X_{East}, Y_{North} and Z_{Down} are the reference coordinates extracted by the SPAN estimation. The PE_{3D} and PE_{2D} are plotted in Figures 5.47 and 5.47 with some meaningful statistics reported in Table 5.4 and 5.5. From the results is possible to state that the 3D positioning accuracy is about 5-7 m considering the 95^{th} percentile, while the 2D positioning accuracy is 5-6 m.

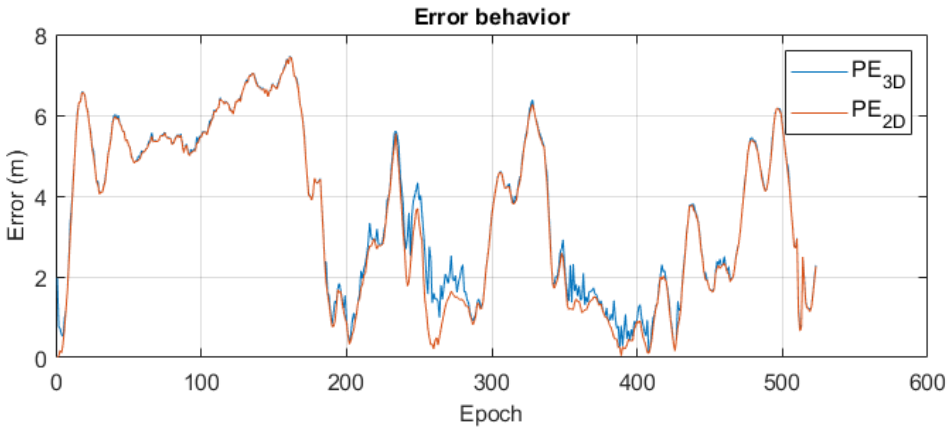


Figure 5.47: 3D and 2D positioning error. Test n°1.

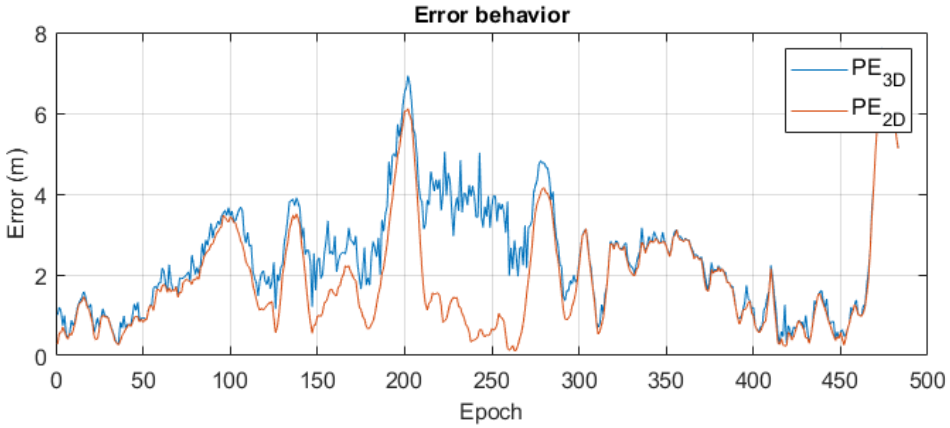


Figure 5.48: 3D and 2D positioning error. Test n°1.

Table 5.4: PE_{3D} statistical characterization for both tests.

Statistics	Test n° 1	Test n° 2
Min [m]	0.101	0.243
Max [m]	7.464	7.262
Mean [m]	3.640	2.510
Std.D. [m]	2.019	1.395
50^{th} precentile	3.788	2.371
70^{th} precentile	5.226	3.029
95^{th} precentile	6.723	5.058

Table 5.5: PE_{2D} statistical characterization for both tests.

Statistics	Test n° 1	Test n° 2
Min [m]	0.0128	0.087
Max [m]	7.450	7.262
Mean [m]	3.488	1.883
Std.D. [m]	2.128	1.342
50^{th} precentile	3.656	1.556
70^{th} precentile	5.206	2.329
95^{th} precentile	6.706	4.691

Chapter 6

CONCLUSIONS

The aims of this doctoral thesis is to investigate on sensors systems and relative methodologies to perform positioning and navigation of people, goods and vehicles in challenging environments and particularly in urban cities. The interest on this research raises from two unsolved problems in the field of navigation:

1. there are not current positioning systems able to ensure the robustness and integrity of the positioning in critical scenarios (indoor, disturbed, shielded, crowded) like the GNSS in outdoor environments;

2. does not exist a positioning and navigation system based on a single technology able to guarantee the continuity of the position estimation in the transition from an open space to an indoor environment, i.e. a seamless navigation system that works every time and everywhere.

These two statements are particularly true when sensors or integrated systems are low-cost, because they are characterized by the introduction of noise in the observations.

To this aims, deep knowledge and study of the state of art related to technologies, methods and algorithms of data fusion is crucial to identify the gap in the current research. What emerged from the literature review has defined the research lines of this thesis. In particular, a wide range of innovative technologies and solutions is observed, based on different physical quantities (both analog and digital) and methodologies. These solutions are used for several applications, therefore numerous requirements must be guaranteed. Moreover, each sensor operates differently according to the the type of environment and provides specific information related to the provided service. Each system has costs, energy consumption, communication infrastructure needs, and other characteristics which highlight their heterogeneity.

In this context, it is clear the need to integrate different technologies to exploit the benefits and to mitigate the limits of each sensor. The multi-sensor hybridization can be considered the first research line of this work.

The second research line, raising from the literature reviews, is the awareness that any solution is strictly environment dependent. Before acquiring data and investigating on positioning algorithm it is mandatory to define which challenges the system should face according to the environment characteristics. This allows to determine the the most suitable sensor or method for the purpose of seamless navigation.

The urban environment is one of the most challenging sites in which perform navigation. It is composed by tall building and threes which mitigate the GNSS signal and generates multipath effects causing erroneous position estimation. It is rich of interference sources for radio-based signals due to the presence of numerous man-made communication infrastructures. Also, it presents indoor and underground spaces where GNSS signals are attenuated or blocked.

All these elements make the research on seamless positioning and navigation in urban environment essential to face the profound changes of our society as regards smart mobility, emergency management and information society.

Urban cities are composed by complex spaces (districts, squares, public offices, private houses, etc.) with different sizes, so that a multi-scale approach is required. In fact, one of the strictest sensor feature is the coverage, i.e. the capability of the system to operate inside an area. This parameter has strongly influenced the selection of sensors in this research. Three spaces have been identified and three different technologies and methods of positioning have been selected. Firstly, with intent to perform positioning in limited spaces, like offices and building floors, the so-called floor scale environment has been defined. In this area, an UWB technology for positioning and tracking has been used and validated.

Secondly, to perform positioning in larger spaces, like hospitals, railway and subway stations, the infrastructure scale has been defined. In this area, image-based positioning approach has been used.

Finally, to perform positioning outdoor and in urban canyons, where the spaces are not restricted, the so-called district scale has been defined and the GNSS positioning with smartphones has been investigated.

The choice of these three methodologies has been further confirmed by empirical data obtained by real applications. For the floor scale, the performances of low-cost UWB systems have been validated in term of real time positioning and ranging. The systems have been used within offices and narrow corridors of the Department of Environment, Land and Infrastructure Engineering of Politecnico di Torino (Italy). The accuracy level reached with this kind of sensor is strongly affected by the environmental noise, nevertheless the error in positioning was of some centimeters.

Regarding the infrastructure scale environment, a research project conducted in 2016 with the aims of perform indoor positioning in an underground metro station, has allowed to investigate on image recognition-based location. In this case, several tests have been performed in Bangbae metro station of Seoul (Republic of Korea). The solution developed by the Geomatic Research Group of the Politecnico di

Torino and validated by the author has demonstrated the possibility to perform real time positioning in man-made crowded environment with a sub-metric level of accuracy and with an high level of reliability. Moreover, such methods allow to estimate not only the position of the imaging sensor but also its orientation.

Finally, the district scale have been considered and the performances of GNSS smartphones chipsets have been assessed. The interest in this technology is increased by the recent possibility to access to the raw observables of the GNSS chipset installed in new-generation mass-market smartphones. From 2016, Google has released a new API for the Android operating systems which allows to acquire directly the GNSS raw data. The test conducted in the Politecnico di Torino (Italy) has shown the capability of these systems to reach high accuracy in the positioning with real time kinematic estimation procedure.

The analysis have been enriched by the statistical characterization of the sensors observation and by focusing on the site-dependent behaviour of the estimates. Moreover, the outcomes of these three case studies demonstrated the systems complementarity and permitted to set up the sensor integration: a multi-sensor platform, composed by UWB transceiver, smartphone device, Imaging sensor and INS.

In this context, a data fusion algorithm has been developed and a theoretical framework has been proposed, taking into account the redundancy of the acquired data and, above all, the main objective of the thesis: providing a seamless navigation solution.

A research project conducted in 2018 by the Finnish Geospatial Research Institute, Helsinki (Finland) gave to the author the chance to investigate in this direction with a real case study. The Kaivos Project consisted in developing and validating low-cost pedestrian navigation solutions for underground mines. The characteristics of mines are very challenging for positioning and navigation procedures. Firstly, they are similar to indoor environments, where the GNSS navigation is not suitable. Then, the presence of ferrometallic minerals can generate interferences in radiofrequency-based technologies. Finally, the low-light conditions represent a strong issue for all the navigation procedures based on images. From this point of view, if a system is able to perform seamless navigation in an underground mine, it is possible to state that it can perform well also in an urban environment.

The above mentioned methodologies used at different urban scales, have been here applied with new-algorithms and filters to obtain a navigation solution. The algorithm, composed by a hybridization of UWB sensor, infrared low-cost camera and inertial platform, has been developed in a loosely coupling approach inside a EKF estimation procedure. The results obtained have been validated in term of accuracy and precision with respect to a reference solution. Unfortunately, the visual odometry algorithm has been applied on a heavily blurred dataset providing no results, although the proposed framework is valid.

The results obtained shows that the UWB/INS integration provide a navigation solution with a mean error of less than 3.6 meters and a standard deviation less

than 2 meters in 3D positioning. In 2D positioning the mean error is less than 3.5 meters with a standard deviation less than 2 meters. Numerous sources of errors have been injected during the test, mainly due to the bad time synchronization and to the various approximations. Nevertheless, the results demonstrate the possibility to use a low-cost integrated system in this critical environment, guaranteeing the indoor-outdoor passage.

A drawback of the proposed solution is the use of UWB sensor in complex and narrow mines. In this case, the economic advantage of low-cost sensors is lost when a capillary infrastructure is needed to obtain the required accuracy.

The main contribution of this work has been the integration of several low-cost sensors in a portable platform for challenging environment navigation. Moreover, the numerous issues identified for each sensor analysed provide a robust basis to upgrade the system for seamless navigation in urban environment. Considering the obtained results, a context awareness procedure should be implemented in order to trigger sensors and weight observations according to the operational context and user behaviour.

6.1 Future activities

In this thesis several positioning solutions have been considered and analyzed in order to cope with the problem of seamless positioning and navigation in challenging environments. Although the performances of different methodologies have been assessed and the benefits of multi-sensor integration have been demonstrated, the proposed solution highlights new research path for future developments of the thesis work.

Regarding the GNSS positioning with smartphones, future activities could concern the use of dual-frequency, multi-constellation receivers which are facing the markets nowadays. The availability of signals from two frequencies allows to correct most of the error sources introduced by ionospheric propagation. At the same time, it opens the way to solve the carrier phase integer ambiguity, enabling RTK and Precise Point Positioning (PPP). Frequency diversity is among the most quoted solutions to increase robustness to interference and jamming.

In the proposed solution UWB systems have been used to provide additional measurements for the data-fusion algorithm. Nevertheless, its data transmission capability could allow to define a geofence, in which a particular response of the positioning algorithm could be triggered. This feature could be particularly interesting to automatically switch between indoor and outdoor solution and to guarantee the seamlessness.

Furthermore, possible future developments of this work, regarding the vision-based navigation, are driven by the newly implementation of machine vision components and the continuing advances of deep learning and artificial intelligence in

embedded systems. With these methods, the limits of popular state-of-art algorithms for pose estimation and feature extraction could be overcome.

Regarding the integration algorithm proposed in the present work, the classical KF techniques could be developed in more outperforming implementations, such as tight and ultra-tight coupling or particle filters. Also artificial intelligence could be investigated in the case where accurate stochastic modeling is not possible.

Bibliography

[1] Sana, "A Survey of Indoor Localization Techniques", *IOSR Journal of Electrical and Electronics Engineering (IOSR-JEEE)*, vol. 6, no. 3, pp. 69–76, 2013.

[2] R. Mautz, "Indoor positioning technologies", PhD thesis, ETH Zurich, Department of Civil, Environmental and Geomatic Engineering, 2012.

[3] L. Faramondi, F. Inderst, F. Pascucci, R. Setola, and U. Delprato, "An enhanced indoor positioning system for first responders", in *2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, IEEE, 2013, pp. 1–8.

[4] J. Torres-Solis, T. H. Falk, and T. Chau, "A review of indoor localization technologies: towards navigational assistance for topographical disorientation", in *Ambient Intelligence*, InTech, 2010.

[5] J. A. P. Montañés, A. M. Rodríguez, and I. S. Prieto, "Smart indoor positioning/location and navigation: A lightweight approach", *International Journal of Interactive Multimedia and Artificial Intelligence (IJIMAI)*, vol. 2, no. 2, pp. 43–50, 2013.

[6] X.-Y. Lin, T.-W. Ho, C.-C. Fang, Z.-S. Yen, B.-J. Yang, and F. Lai, "A mobile indoor positioning system based on iBeacon technology", in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2015, pp. 4970–4973.

[7] P. Jordil, S. Knut, B. Pettersson, and K. Schneider, "Method of using multi-axis positioning and measuring system", U.S. Patent 7 676 942.

[8] A. Fink, H. Beikirch, M. Voss, and C. Schroder, "RSSI-based indoor positioning using diversity and inertial navigation", in *2010 International Conference onIndoor Positioning and Indoor Navigation (IPIN)*, IEEE, 2010, pp. 1–7.

[9] P. Dabove and V. Di Pietra, "Towards high accuracy GNSS real-time positioning with smartphones", *Advences in Space Research*, pp. 1–9, 2018.

[10] F. Chiabrando, V. Di Pietra, A. M. Lingua, Y. Cho, and J. Jeon, "An original application of image recognition based location in complex indoor environments", *ISPRS International Journal of Geo-Information*, vol. 6(2), 2017.

[11] V. Di Pietra, M. Piras, P. Dabove, A. Abdul Jhabbar, and A. Syed Kazim, "Indoor positioning con tecniche Ultra Wide Band: funzionamento, test e risultati", in *21a Conferenza Nazionale ASITA*, Asita, 2017, pp. 443–445.

[12] F. Chiabrando, V. Di Pietra, A. M. Lingua, J. Jeon, and Y. Cho, "Solid Image generation for indoor image based positioning. The Bangbae metro station in Seoul.", in *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Danvers: IEEE, 2016.

[13] P. Dabove, V. Di Pietra, M. Piras, A. A. Jabbar, and S. A. Kazim, "Indoor positioning using Ultra-wide band (UWB) technologies: Positioning accuracies and sensors' performances", in *2018 IEEE/ION Position, Location and Navigation Symposium, PLANS 2018 - Proceedings*, Institute of Electrical and Electronics Engineers Inc., 2018, pp. 175–184.

[14] P. Dabove, V. D. Pietra, and A. M. Lingua, "Positioning Techniques with Smartphone Technology: Performances and Methodologies in Outdoor and Indoor Scenarios", in *Smartphones from an Applied Research Perspective*. InTech, 2017, pp. 163–183.

[15] P. D. Groves, *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems, Second Edition*. Artech House, Apr. 2013, ISBN: 978-1-60807-005-3.

[16] R. Mautz, "Indoor positioning–an ad-hoc positioning system", *Geodezija ir Kartografija*, vol. 34, no. 2, pp. 66–70, 2008.

[17] G. Deak, K. Curran, and J. Condell, "A survey of active and passive indoor localisation systems", *Computer Communications*, vol. 35, no. 16, pp. 1939–1954, 2012.

[18] J. J. Liu, C. Phillips, and K. Daniilidis, "Video-based localization without 3D mapping for the visually impaired", in *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, 2010, pp. 23–30.

[19] J. Z. Liang, N. Corso, E. Turner, and A. Zakhor, "Reduced-complexity data acquisition system for image-based localization in indoor environments", in *2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, IEEE, 2013, pp. 1–9.

[20] Q. Anwar, A. W. Malik, and B. Thornberg, "Design of coded reference labels for indoor optical navigation using monocular camera", in *2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, IEEE, 2013, pp. 1–8.

[21] Z. Zhou, X. Chen, Y.-C. Chung, Z. He, T. X. Han, and J. M. Keller, "Video-based activity monitoring for indoor environments", in *2009 IEEE International Symposium on Circuits and Systems (ISCAS)*, IEEE, 2009, pp. 1449–1452.

[22] R. Want, D. Goldberg, A. G. LaMarca, T. A. Cass, and M. D. Weiser, *Infrared beacon positioning system*, U.S. Patent, Apr. 2001.

[23] W.-Y. Chung *et al.*, "Enhanced RSSI-based real-time user location tracking system for indoor and outdoor environments", in *2007 International Conference on Convergence Information Technology,*, IEEE, 2007, pp. 1213–1218.

[24] R. E. Kalman, "A New Approach to Linear Filtering and Prediction Problems", *Journal of Basic Engineering*, vol. 82, no. 1, pp. 35–45, Mar. 1960.

[25] M. G. Petovello, K. O'Keefe, G. Lachapelle, and M. E. Cannon, "Consideration of time-correlated errors in a Kalman filter applicable to GNSS", *Journal of Geodesy*, vol. 83, no. 1, pp. 51–56, 2009.

[26] C. Hide, T. Moore, and M. Smith, "Adaptive Kalman filtering for low-cost INS/GPS", *The Journal of Navigation*, vol. 56, no. 1, pp. 143–152, 2003.

[27] G. Forlani, "Metodi robusti di stima in geodesia e fotogrammetria", 1990, pp. 124–311.

[28] M. Riani, A. Cerioli, A. C. Atkinson, and D. Perrotta, "Monitoring robust regression", *Electronic Journal of Statistics*, vol. 8, no. 1, pp. 646–677, 2014.

[29] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", *Commununications of the ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981.

[30] F. C. C. (2002b), *Revision of Part 15 of the Commission's Rules Regarding Ultra-Wideband Transmission Systems*, Apr. 2002. [Online]. Available: https://www.fcc.gov/document/revision-part-15-commissions-rules-regarding-ultra-wideband (visited on 12/28/2018).

[31] R. I. Hartley and J. L. Mundy, "Relationship between photogrammmetry and computer vision", in *Integrating photogrammetric techniques with scene analysis and machine vision*, International Society for Optics and Photonics, vol. 1944, 1993, pp. 92–106.

[32] D. H. Hubel and T. N. Wiesel, "Effects of monocular deprivation in kittens", *Naunyn-Schmiedebergs Archiv for Experimentelle Pathologie und Pharmakologie*, vol. 248, no. 6, pp. 492–497, 1964.

[33] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[34] R. Hartley and A. Zisserman, *Multiple view geometry in computer vision*. Cambridge university press, 2003.

[35] M. V. Srinivasan, J. S. Chahl, K. Weber, S. Venkatesh, M. G. Nagle, and S.-W. Zhang, "Robot navigation inspired by principles of insect vision", *Robotics and Autonomous Systems*, vol. 26, no. 2-3, pp. 203–216, 1999.

[36] J. J. Moré, "The Levenberg-Marquardt algorithm: implementation and theory", in *Numerical analysis*, Springer, 1978, pp. 105–116.

[37] O. D. Faugeras and F. Lustman, "Motion and structure from motion in a piecewise planar environment", *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 2, no. 03, pp. 485–508, 1988.

[38] E. Malis and M. Vargas, "Deeper understanding of the homography decomposition for vision-based control", PhD thesis, INRIA, 2007.

[39] M. Wadenback, "Homography-Based Positioning and Planar Motion Recovery.", PhD thesis, Lund University, Sweden, 2017.

[40] D. Scaramuzza and F. Fraundorfer, "Visual odometry [tutorial]", *IEEE robotics & automation magazine*, vol. 18, no. 4, pp. 80–92, 2011.

[41] H. Durrant-Whyte and T. Bailey, "Simultaneous localisation and mapping (SLAM): Part I", *IEEE robotics & automation magazine*, vol. 13, no. 2, pp. 99–110, 2006.

[42] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation", in *2007 IEEE international conference on Robotics and automation*, IEEE, 2007, pp. 3565–3572.

[43] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf: Speeded up robust features", in *European conference on computer vision*, Springer, 2006, pp. 404–417.

[44] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF", in *2011 IEEE international conference on Computer Vision (ICCV)*, IEEE, 2011, pp. 2564–2571.

[45] J. C. McGlone, E. M. Mikhail, J. S. Bethel, and R. Mullen, "Manual of photogrammetry", American society for photogrammetry and remote sensing Bethesda, MD, 2004.

[46] *Digital in 2018 Report: gli utenti Internet nel mondo superano i 4 miliardi. In Italia sono più di 43 milioni*, Jan. 2018. [Online]. Available: https://wearesocial.com/it/blog/2018/01/global-digital-report-2018 (visited on 11/26/2018).

[47] E. Kaplan and C. Hegarty, *Understanding GPS: principles and applications*. Artech House, 2005.

[48] P. Misra and P. Enge, *Global Positioning System: signals, measurements and performance second edition*. Massachusetts: Ganga-Jamuna Press, 2006.

[49] B. Hofmann-Wellenhof, H. Lichtenegger, and E. Wasle, *GNSS–global navigation satellite systems: GPS, GLONASS, Galileo, and more*. Springer Science & Business Media, 2007.

[50] P. Dabove, A. M. Manzino, and C. Taglioretti, "GNSS network products for post-processing positioning: limitations and peculiarities", *Applied geomatics*, vol. 6, no. 1, pp. 27–36, 2014.

[51] A. Cina, P. Dabove, A. Manzino, and M. Piras, "Augmented positioning with CORSs network services using GNSS mass-market receivers", in *2014 IEEE/ION Position, Location and Navigation Symposium - PLANS 2014*, IEEE, 2014, pp. 359–366.

[52] F. S. T. Van Diggelen, *A-GPS: Assisted GPS, GNSS, and SBAS*. Artech House, 2009.

[53] *Raw GNSS Measurements*, Jan. 2018. [Online]. Available: https://developer.android.com/guide/topics/sensors/gnss (visited on 10/20/2018).

[54] *Android.location Documentation*, Jan. 2018. [Online]. Available: https://developer.android.com/reference/android/location/package-summary (visited on 10/20/2018).

[55] D. Gebre-Egziabher and S. Gleason, *GNSS applications and methods*. Artech House, 2009.

[56] M. El-Diasty and S. Pagiatakis, "Calibration and stochastic modelling of inertial navigation sensor errors", *Journal of Global Positioning Systems*, vol. 7, no. 2, pp. 170–182, 2008.

[57] A. Solimeno, "Low-cost INS/GPS data fusion with extended Kalman filter for airborne applications", *Masters of Science, Universidade Technica de Lisboa*, 2007.

[58] M. G. Petovello, *Real-time integration of a tactical-grade IMU and GPS for high-accuracy positioning and navigation*. Citeseer, 2003.

[59] C. Brown, *Real-time location of jena's buses and trams with ubisense RTLS*, 2010.

[60] *Ubisense launches Intrinsically Safe location tracking tags for personnel safety in the Oil and Gas industry*, Jan. 2010. [Online]. Available: https://www.cambridgenetwork.co.uk/news/ubisense-launches-is-location-tracking-tags-for-personnel/ (visited on 03/27/2019).

[61] M. Baum, "RTL in Longueuil selects bus yard management solution provided by Solotech", *ISR Transit and Ubisense*, 2011.

[62] G. Cheng, "Accurate TOA-based UWB localization system in coal mine based on WSN", *Physics Procedia*, vol. 24, pp. 534–540, 2012.

[63] L. Jiang, L. N. Hoe, and L. L. Loon, "Integrated UWB and GPS location sensing system in hospital environment", in *2010 5th IEEE Conference on Industrial Electronics and Applications*, IEEE, 2010, pp. 286–289.

[64] J. Chóliz, M. Eguizabal, Á. Hernández-Solana, and A. Valdovinos, "Comparison of algorithms for uwb indoor location and tracking systems", in *2011 IEEE 73rd Vehicular Technology Conference (VTC Spring)*, IEEE, 2011, pp. 1–5.

[65] E. García, P. Poudereux, Á. Hernández, J. Ureña, and D. Gualda, "A robust UWB indoor positioning system for highly complex environments", in *2015 IEEE International Conference on Industrial Technology (ICIT)*, IEEE, 2015, pp. 3386–3391.

[66] L. Deambrogio, C. Palestini, F. Bastia, G. Gabelli, G. E. Corazza, and J. Samson, "Impact of high-end receivers in a peer-to-peer cooperative localization system", in *2010 Ubiquitous Positioning Indoor Navigation and Location Based Service*, IEEE, 2010, pp. 1–7.

[67] M.-F. Tsai, P.-C. Wang, C.-K. Shieh, W.-S. Hwang, N. Chilamkurti, S. Rho, and Y. S. Lee, "Improving positioning accuracy for VANET in real city environments", *The Journal of Supercomputing*, vol. 71, no. 6, pp. 1975–1995, 2015.

[68] K. Liu and H. B. Lim, "Positioning accuracy improvement via distributed location estimate in cooperative vehicular networks", in *2012 15th International IEEE Conference on Intelligent Transportation Systems*, IEEE, 2012, pp. 1549–1554.

[69] *TREK1000 Evaluation Kit*, 2018. [Online]. Available: https://www.decawave.com/product/trek1000-evaluation-kit/ (visited on 10/20/2018).

[70] *Standard Evaluation Kit with BeSpoon Protocol*, 2018. [Online]. Available: http://bespoon.com/shop/en/home/13-standart-evaluation-kit-with-bespoon-protocol.html (visited on 10/20/2018).

[71] *Ubisense Dimension4*, 2018. [Online]. Available: https://www.ubisense.net/product/dimension4 (visited on 10/20/2018).

[72] *Utra Wideband (UWB) Technology*, 2018. [Online]. Available: https://www.zebra.com/gb/en/products/location-technologies/ultra-wideband.html (visited on 10/20/2018).

[73] G. Ghinamo, C. Corbi, G. Francini, S. Lepsoy, P. Lovisolo, A. Lingua, and I. Aicardi, "The MPEG7 Visual Search Solution for image recognition based positioning using 3D models", in *Proceedings of the 27th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2014)*, p. 12.

[74] T. Sikora, "The MPEG-7 visual standard for content description-an overview", *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 696–702, Jun. 2001.

[75] 14:00-17:00, *ISO/IEC 15938-13:2015*. [Online]. Available: http://www.iso.org/cms/render/live/en/sites/isoorg/contents/data/standard/06/53/65393.html.

[76] S. Lepsoy, G. Francini, G. Cordara, and P. P. B. d. Gusmao, "Statistical modelling of outliers for fast visual search", in *2011 IEEE International Conference on Multimedia and Expo*, Jul. 2011, pp. 1–6.

[77] C. Giovanni, F. Gianluca, L. Skjalg, and D. G. Pedro Porto Buarque, "Method and system for comparing images", U.S. Patent US13981824.

[78] R. Mullen, *Manual of Photogrammetry. Fifth Edition*, English, 5 edition. Bethesda, Md: ASPRS, 2004.

[79] H. Karara, "Non-Topographic Photogrammetry, American Society for Photogrammetry and Remote Sensing", *Science and Engineering Series, Virginia, USA*, 1989.

[80] W. Baarda, "A testing procedure for use in geodetic networks.", 1968.

[81] K. Kraus, *Photogrammetry: geometry from images and laser scans*. Walter de Gruyter, 2011.

[82] C. Balletti, F. Guerra, P. Vernier, N. Studnicka, J. Riegl, S. Orlandini, and M. Srl, "Practical comparative evaluation of an integrated hybrid sensor based on Photogrammetry and Laser Scanning for Architectural Representation", in *ISPRS, International Archives of Photogrammetry and Remote Sensing", Commission V, ISPRS XX Congress, Istanbul, Turchia*, 2004, pp. 12–23.

[83] H. Men, B. Gebre, and K. Pochiraju, "Color point cloud registration with 4D ICP algorithm", in *2011 IEEE International Conference on Robotics and Automation (ICRA)*, IEEE, 2011, pp. 1511–1516.

[84] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes", in *Sensor Fusion IV: Control Paradigms and Data Structures*, International Society for Optics and Photonics, vol. 1611, 1992, pp. 586–607.

[85] K. M. J. Pesyna, R. W. J. Heath, and T. E. Humphreys, "Centimeter Positioning with a Smartphone-Quality GNSS Antenna", eng, Sep. 2014. (visited on 02/15/2019).

[86] X. Zhang, X. Tao, F. Zhu, X. Shi, and F. Wang, "Quality assessment of GNSS observations from an Android N smartphone and positioning performance analysis using time-differenced filtering approach", en, *GPS Solutions*, vol. 22, no. 3, p. 70, May 2018, ISSN: 1521-1886. (visited on 02/11/2019).

[87] D. Laurichesse, C. Rouch, F.-X. Marmet, and M. Pascaud, "Smartphone Applications for Precise Point Positioning", Portland, Oregon, Nov. 2017, pp. 171–187. (visited on 02/15/2019).

[88] E. Realini, S. Caldera, L. Pertusini, and D. Sampietro, "Precise GNSS Positioning Using Smart Devices", en, *Sensors*, vol. 17, no. 10, p. 2434, Oct. 2017.

[89] *Rete GNSS Interregionale Piemonte-Lombardia*. [Online]. Available: https://www.spingnss.it/spiderweb/frmIndex.aspx (visited on 01/23/2019).

[90] P. J. Teunissen, "The least-squares ambiguity decorrelation adjustment: a method for fast GPS integer ambiguity estimation", *Journal of geodesy*, vol. 70, no. 1-2, pp. 65–82, 1995.

[91] X.-W. Chang, X. Yang, and T. Zhou, "MLAMBDA: a modified LAMBDA method for integer least-squares estimation", *Journal of Geodesy*, vol. 79, no. 9, pp. 552–565, 2005.

[92] M. of Employment and the economy, *Making Finland a leader in the sustainable extractive industry - action plan*, 2013. [Online]. Available: https://tem.fi/documents/1410877/3437254/Making+Finland+a+leader+in+the+sustainable+extractive+industry+04072013.pdf (visited on 10/20/2018).

[93] *Read "Coal: Research and Development to Support National Energy Policy" at NAP.edu*, en. DOI: 10.17226/11977. [Online]. Available: https://www.nap.edu/read/11977/chapter/6 (visited on 01/15/2019).

[94] S. Walters and T. Kojovic, "Geometallurgical mapping and mine modelling (GEMIII)-the way of the future", in *SAG 2006*, vol. 4, 2006, pp. 411–425.

[95] *Kaivosmine Project*, en, Jun. 2016. [Online]. Available: https://kaivosmine.net/inenglish/ (visited on 01/15/2019).

[96] S. Kaasalainen, S. GrÖhn, O. Nevalainen, and T. Hakala, "Work in Progress: Combining Indoor Positioning and 3D Point Clouds from Multispectral Lidar", *International Conference on Indoor Positioning and Indoor Navigation*, p. 4, 2016.

[97] S. Kaasalainen, L. Ruotsalainen, M. Kirkko-Jaakkola, O. Nevalainen, and T. Hakala, "Towards multispectral, multi-sensor indoor positioning and target identification", *Electronics Letters*, vol. 53, no. 15, pp. 1008–1011, 2017.

[98] L. Ruotsalainen, M. Kirkko-Jaakkola, J. Rantanen, and M. Mäkelä, "Error Modelling for Multi-Sensor Measurements in Infrastructure-Free Indoor Navigation", *Sensors*, vol. 18, no. 2, Feb. 2018. (visited on 12/06/2018).

[99] R. Mautz and S. Tilch, "Survey of optical indoor positioning systems", in *2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, IEEE, 2011, pp. 1–7.

[100] S.-H. Jung, J. Eledath, S. Johansson, and V. Mathevon, "Egomotion estimation in monocular infra-red image sequence for night vision applications", in *2007 IEEE Workshop on Applications of Computer Vision (WACV'07)*, IEEE, 2007, pp. 8–8.

[101] J. Poujol, C. A. Aguilera, E. Danos, B. X. Vintimilla, R. Toledo, and A. D. Sappa, "A Visible-Thermal Fusion Based Monocular Visual Odometry", in *Robot 2015: Second Iberian Robotics Conference*, Springer, 2016, pp. 517–528.

[102] *MTi-G-710 - Products - Xsens 3D motion tracking*. [Online]. Available: https://www.xsens.com/products/mti-g-710/ (visited on 01/15/2019).

[103] J. Wong, D. Sang, and C.-S. Peng, "An Android Geofencing App for Autonomous Remote Switch Control", *International Journal of Computer and Information Engineering*, vol. 11, no. 3, p. 9, 2017.

[104] *An extremely effective and outstanding Matlab algorithm for recovering image from motion blur*, May 2016. [Online]. Available: https://github.com/Thelordofdream/Recover-image-from-motion-blur (visited on 01/15/2019).

[105] P. Aggarwal, Z. Syed, X. Niu, and N. El-Sheimy, "A standard testing and calibration procedure for low cost MEMS inertial sensors and units", *The Journal of Navigation*, vol. 61, no. 2, pp. 323–336, 2008.

[106] P. J. Teunissen, "The non-linear 2D symmetric Helmert transformation: an exact non-linear least-squares solution", *Bulletin geodesique*, vol. 62, no. 1, pp. 1–16, 1988.

[107] ETSI, *TS 103 246-5: Satellite Earth Stations and Systems (SES), GNSS based location systems, Performance Test Specification*.

[108] T. E. Humphreys, M. Murrian, F. v. Diggelen, S. Podshivalov, and K. M. Pesyna, "On the feasibility of cm-accurate positioning via a smartphone's antenna and GNSS chip", in *2016 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, Apr. 2016, pp. 232–242.

This Ph.D. thesis has been typeset by means of the TEX-system facilities. The typesetting engine was pdfLATEX. The document class was `toptesi`, by Claudio Beccari, with option `tipotesi=scudo`. This class is available in every up-to-date and complete TEX-system installation.